

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES**  
**Departamento de Economía Financiera y Contabilidad III**



**TESIS DOCTORAL**

**Técnicas avanzadas de análisis financiero. Aplicación al diagnóstico y pronóstico en oncología clínica**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Carlos Humberto Ruiz Freire**

Directores

**Juan Antonio Maroto Acín**  
**Beatriz Pérez-Villamil Salgado**  
**Fernando Martín Sánchez**

**Madrid, 2017**

**TÉCNICAS AVANZADAS DE ANÁLISIS  
FINANCIERO. APLICACIÓN AL DIAGNÓSTICO Y  
PRONÓSTICO EN ONCOLOGÍA CLÍNICA**

**Carlos Humberto Ruiz Freire**

**Directores de la Tesis**

**Juan Antonio Maroto Acín**

**Beatriz Pérez-Villamil Salgado**

**Fernando Martín Sánchez**

**Facultad de Ciencias Económicas y Empresariales  
Departamento de Economía Financiera y Contabilidad III  
Universidad Complutense**

**2015**



A mis padres Carlos Humberto y María José

A la Memoria de mis abuelos, Martín Freire Alpañil y María de Los  
Ángeles Pérez Martínez



## **AGRADECIMIENTOS**

La presente Tesis es un esfuerzo en el cual, directa o indirectamente, participaron varias personas leyendo, opinando, corrigiendo, teniendo paciencia, dando ánimo, acompañando en los momentos de crisis y en los momentos de felicidad.

Especialmente quisiera mostrar mi gratitud y más sincero cariño a los directores de esta tesis: Dr. Juan Antonio Maroto Acín, Dra Beatriz Pérez-Villamil Salgado y el Dr. Fernando Martín Sánchez por la orientación, el seguimiento y la supervisión continua de la misma, pero sobre todo por la motivación y el apoyo recibido a lo largo de estos años.

Al Dr. Juan Antonio Maroto me gustaría agradecerle su exquisito trato, el tiempo invertido, su enorme paciencia, su sensatez y pragmatismo con el que ha abordado las distintas etapas de la investigación y que su puerta siempre ha estado abierta, tanto para temas académicos como para los no académicos. De idéntica forma me gustaría agradecerle a la Dra. Beatriz Pérez su plena accesibilidad en todo momento para resolver las más variopintas cuestiones que han surgido a pie de laboratorio. Nunca ha escatimado esfuerzos para adentrarse en conceptos económicos para poder explicarme aspectos biológicos.

En esta misma línea, resulta del todo necesario señalar que este proyecto no hubiera nacido sin la orientación de la Dra. María Isabel Martínez

Torre-Enciso quien me animó y apoyó en iniciar esta andadura y siempre ha sabido encontrar el tiempo y palabras adecuadas para guiarme en estos años.

Quisiera hacer extensiva mi gratitud a D. Jesús Aparicio por su inestimable colaboración en la revisión matemática del modelo, por el apoyo ofrecido, por escucharme, aconsejarme siempre y, sobre todo, por esa gran amistad que me brindó y me brinda.

Todo esto nunca hubiera sido posible sin el amparo incondicional de mi familia: mis padres, mi mujer y mi hermano. Por eso me gustaría expresarles de una manera muy especial mi más profundo agradecimiento por su comprensión y paciencia infinita, apoyo diario y ánimo recibido durante todos los años que ha durado este viaje.

A todos ellos, muchas gracias.

## ABSTRACT

The healing of a neoplastic process is not always possible, so during the last decades oncological studies have been seeking for a better understanding of the nature of cancer that allows an increase in patient survival. These studies were performed in order to follow more refined surgical techniques, more effective drugs with fewer adverse effects, etc.

Technological progress has contributed in recent years to deepen molecular studies of cancer, allowing genes by filtering and subsequent clustering, identify predictive biomarkers that translate into customized strategies for numerous subtypes of cancers.

The purpose of this thesis is to provide a theoretical framework new to the field of molecular oncology, which identifies target genes. It is a new proposal based on the expression profiles<sup>1</sup> obtained in terms of a new parameter to classify tumors by histopathology, based on the main classifications such as: Dukes<sup>2</sup>, and Coller and TNM Astler as well as the variability in expression of genes. Perhaps this new approach can be a contribution in the near future to help explain the reasons why the same treatment in patients with the same tumor stage generates different responses in different patients.

---

<sup>1</sup> Value of the expression of a gene that gives us the microarray.

<sup>2</sup> Published in 1932. For more information see Annex IV.

To address this question, one of the parameters was used in the characterization of market risk: the beta of the beta of a financial asset. This concept is transferred to molecular oncology. Gene identification is carried out by taking into account the expression profiles obtained according to the beta genes.

The study includes 96 samples, 89 tumor samples and 7 normal or control samples. The data has been transferred by the Department of Molecular Oncology at the Clinical Hospital of Madrid between 2001 and 2006. Patients were selected so that all stages<sup>3</sup> of tumor development were present in the study (24 in stage A, 26 in stage B, 19 stage C and 19 stage D), more or less evenly.

Possible implications of improved methods of identifying genes in the healthcare field are many, among which we can highlight:

1. Improvement of personalized treatments for patients.
2. Advances in the design of new and more specific drugs with less toxicity than the existing drugs, improving the quality of life of cancer patients.
3. Improved cancer prevention.
4. Increase in the hope of survival of cancer patients.

Of the 14764 genes, 521 have a beta greater than 1.4 and are out of range (Median – 1.7 standard deviations, Medium + 1.7 standard deviations) in at least more than 8 samples. This set of genes represents 3.5288% of the total

---

<sup>3</sup> The aforementioned stages correlate with the DUKES classification.

gene. If the variability of a sample is partitioned equally between all genes of that sample; that is, if all genes from a sample have provided the same variability of the sample, the value which will provide the set of genes to the total variability of the samples should be from 3.52% and, as already noted, does not. The variability that this set of 521 genes selected contributes to tumor and normal samples is 5.7594%, and 1.1642% respectively. The variability of these genes contributing to tumor samples is 1.6320 times more than would correspond if the variability is partitioned equally among all the genes. In view of these data, it seems clear the importance of these 521 genes.

Looking more closely at the data variability it has been found that this group of genes contributes to tumor samples 1.6421% versus 3.5288%, which would be expected. That is, they account for less than half of variability than if the contribution of each gene were the same, in the normal samples, namely 0.4653 times less.

Comparing the variability contributed by the 521 genes in tumor samples compared to normal samples is observed to be 3.5072 times, more than triple.

It has been taken as a hypothesis that the variability is not directly related to tumor malignancy. Such that, the less variability present a sample, the less malignant the sample is expected to be. This hypothesis is verified both by ordering samples depending on the variability of their genes and the subsequent hierarchical clustering of the samples. Samples which exhibit less variability are precisely normal samples.

The 14764 genes explain the variability of the whole of the 96 samples and provide a sort based variability. If, in a selection of genes, they are able to order samples similarly, meaning you could work with this new subset of genes since we would be providing the same information as the original input data.

To reorder the samples only genes whose beta is greater than one point six and at least 8 samples over the selected gene is outside the range, management remains very similar.

The ranking obtained from the model, as can be seen in the image below:

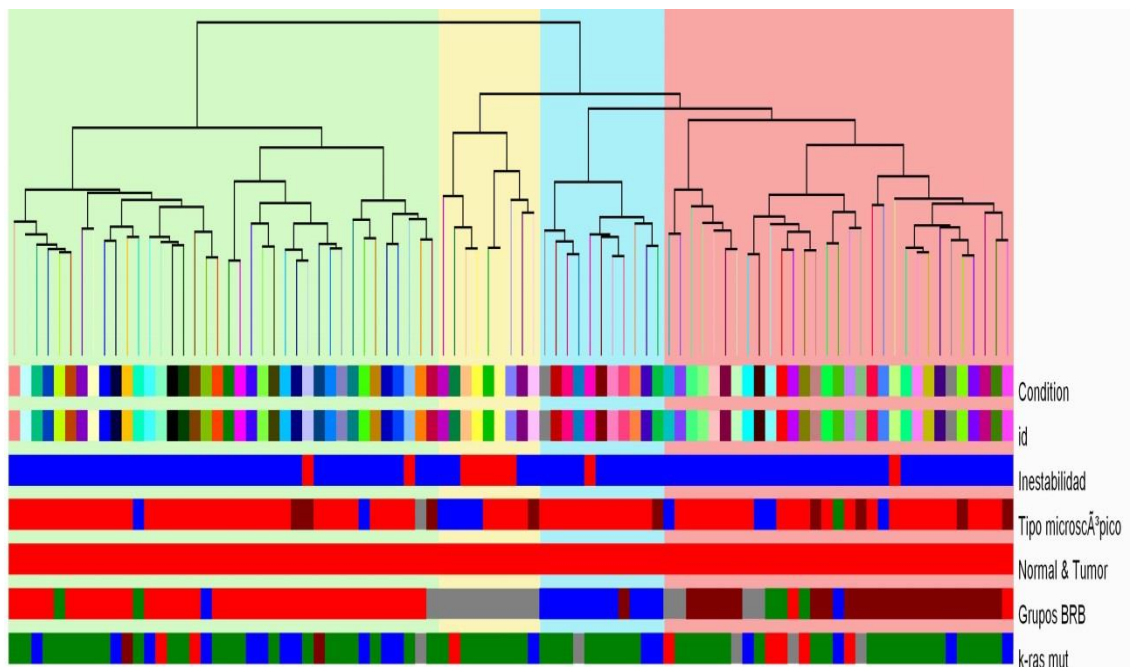


Figure 1: Graph clustering performed.

Source: Author's own.

The green cluster completely classifies in subgroup 1 colorectal cancer, having low levels of stroma. The blue cluster classifies the subtype 2 colorectal cancer associated with immunoglobulins. The red cluster classifies subtype 3 colorectal cancer that has a high stroma. Finally, the yellow cluster classifies the subtype 4 colon cancer unstable mucinous

Therefore, one can conclude that the technique used in this project that selected 521 genes out of the initial 14,764 initial classifies under beta parameter is valid for the 96 samples analyzed.

## RESUMEN

La curación de un proceso neoplásico no siempre es posible, por lo que durante las últimas décadas los estudios oncológicos realizados buscaban una mejor comprensión de la naturaleza del cáncer que permitiese obtener para los pacientes supervivencias cada vez mayores. Estos estudios se realizaban con el fin de conseguir técnicas quirúrgicas más depuradas, fármacos más eficaces y con menos efectos adversos, etc.

El propósito de esta Tesis Doctoral es ofrecer un marco teórico, novedoso en el campo de la oncología molecular, que haga posible la identificación de genes diana. Se desea aportar una nueva propuesta basándose en los perfiles de expresión<sup>4</sup> obtenidos en función de un nuevo parámetro que clasifique los tumores no por su histopatología, fundamento en el que se sustentan las principales clasificaciones tumorales como por ejemplo: Dukes<sup>5</sup>, Astler y Coller o TNM, sino por la variabilidad en la expresión de los genes.

Para abordar esta cuestión, se utiliza uno de los parámetros empleados en la caracterización del riesgo bursátil: la beta de un activo financiero. Se traslada este concepto a la oncología molecular. La identificación de genes se realiza teniendo en cuenta los perfiles de expresión obtenidos en función de las betas de los genes.

El estudio realizado contempla 96 muestras, 89 tumorales y 7 de normales o de control. Los datos han sido cedidos por el Departamento de

---

<sup>4</sup> Valor de la expresión de un gen que nos facilita el microarray.

<sup>5</sup> Publicada en 1932. Para más información véase el Anexo IV



Oncología Molecular del Hospital Clínico de Madrid entre el año 2001 y el 2006. Los pacientes han sido seleccionados de manera que todos los estadios<sup>6</sup> del tumor estuvieran presentes en el estudio (24 en el estadio A, 26 en el estadio B, 19 en el estadio C y 19 en el estadio D), más o menos de manera homogénea.

Las posibles implicaciones que tendría una mejora en los métodos de identificación de genes en el campo sanitario son muchísimas, de entre las que se pueden destacar:

1. Mejora de los tratamientos personalizados para los pacientes.
2. Avance en el diseño de nuevos fármacos mucho más específicos y con menos toxicidad que los existentes mejorando la calidad de vida de los pacientes con cáncer.
3. Mejora en la prevención del cáncer.
4. Aumento de la esperanza de supervivencia de los pacientes oncológicos.

De los 14764 genes, 521 tienen una beta mayor a 1,4 y están fuera del intervalo (Mediana – una coma siete desviaciones típicas, Mediana + una coma siete desviaciones típicas), en al menos, más de 8 muestras. Este conjunto de genes representa un 3.5288% respecto al total de genes. Si la variabilidad de una muestra se repartiera alícuotamente entre todos los genes de esa muestra; es decir, si todos los genes de una muestra tienen aportan la misma variabilidad a dicha muestra, el valor que aportara este conjunto de genes a la

---

<sup>6</sup> Los estadios mencionados corresponden con la clasificación DUKES.

variabilidad total de las muestras debería ser del 3.52% y, como se ha constatado, esto no ocurre. La variabilidad que este conjunto de 521 genes seleccionados aporta a las muestras tumorales y normales es de 5,7594%, y 1,1642% respectivamente. La variabilidad que aportan estos genes a las muestras tumorales es 1,6320 veces más de lo que le correspondería si la variabilidad se repartiera de forma alícuota entre todos los genes. A la vista de estos datos, parece clara la importancia de estos 521 genes.

Al analizar más detenidamente los datos se comprueba que la variabilidad que este grupo de genes aporta a las muestras normales es del 1,6421% frente al 3,5288% que cabría esperar. Es decir, explican menos de la mitad de variabilidad de lo que, si la contribución de cada gen fuera la misma, en las muestras normales, concretamente 0.4653 veces menos.

Al comparar la variabilidad que aportan los 521 genes en las muestras tumorales frente a las muestras normales se observa que es 3,5072 veces, más del triple.

Se ha tomado como hipótesis de partida que la no variabilidad tiene una relación directa con la malignidad del tumor. De tal manera que, cuanto menos variabilidad presente una muestra, menos malignidad cabe esperar que tenga dicha muestra. Esta hipótesis se verifica tanto al ordenar las muestras en función de la variabilidad de sus genes como en la posterior agrupación jerárquica de las muestras. Las muestras que menos variabilidad presentan son, precisamente, las muestras normales.

Los 14764 explican toda la variabilidad de cada una de las 96 muestras y establecen una ordenación en función de su variabilidad. Si al realizar una

selección de genes, éstos son capaces de ordenar a las muestras de manera similar, significaría que podría trabajar con este nuevo subconjunto de genes puesto que estarían aportando la misma información que los datos originales de partida.

Al ordenar nuevamente las muestras únicamente con los genes cuya beta sea mayor a uno coma seis y en, al menos, más de 8 muestras el gen esté fuera del intervalo seleccionado, la ordenación se mantiene muy similar.

La clasificación jerárquica que se obtiene del modelo, tal y como puede observarse en la siguiente imagen, es:

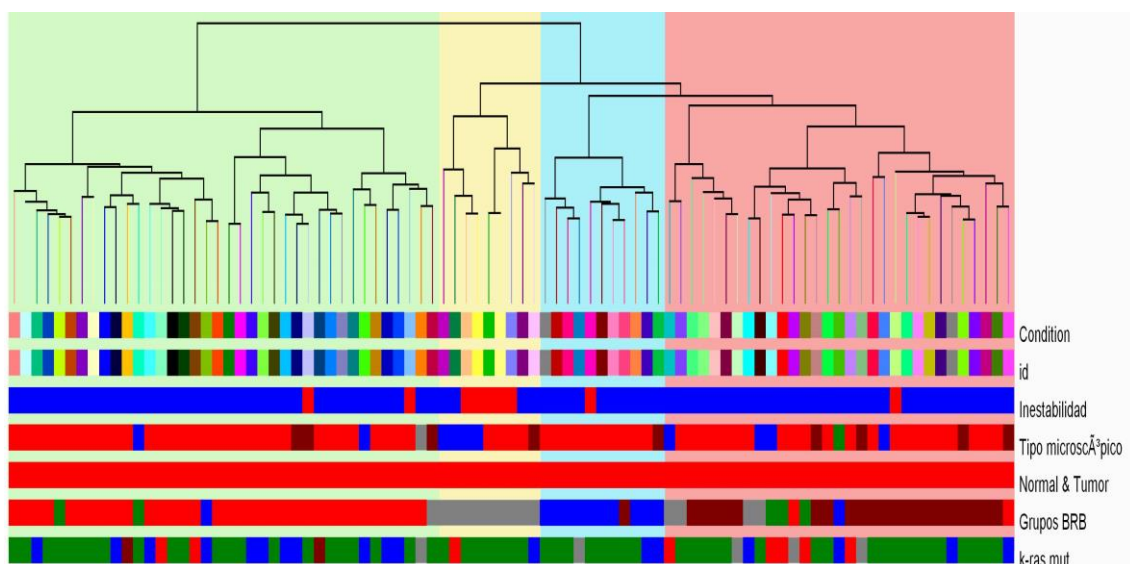


Imagen 1: Gráfico del clustering efectuado.

Fuente: elaboración propia.

El clúster de color verde clasifica por completo al subgrupo 1 de cáncer colorrectal, el que presenta bajos niveles de estroma. El clúster de color azul clasifica al subtipo 2, cáncer colorrectal relacionado con las inmunoglobinas. El clúster de color rojo clasifica al subtipo 3, cáncer colorrectal que presenta alto

en de estroma.Finalmente, el clúster de color amarillo clasifica al subtipo 4, cáncer de colon inestable mucinoso

Por todo ello, se puede concluir que la técnica empleada en este trabajo que selecciona 521 genes de los 14764 iniciales y los clasifica en virtud del parámetro beta, es válida para las 96 muestras analizadas.

## CONTENIDOS

AGRADECIMIENTOS .....	I
ABSTRACT .....	VIII
CAPÍTULO I. INTRODUCCIÓN.....	1
1. 1. INTRODUCCIÓN .....	1
1.2. OBJETIVOS.....	3
1.3. JUSTIFICACIÓN .....	4
1.4. LIMITACIONES .....	19
CAPÍTULO II. FUNDAMENTOS CONCEPTUALES.....	20
2.1. REVISIÓN DE LOS PRINCIPALES MODELOS DE LA TEORÍA DE CARTERAS.....	20
2.1.1. Markowitz .....	20
2.1.2. CAPM.....	36
2.1.3. Modelo de Mercado.....	47
2.1.4. La Modificación de Treynor .....	60
2.2. APLICACIÓN A LA ONCOLOGÍA .....	63
2.2.1. Expresión de un gen .....	63
2.2.2. Riesgo de un gen .....	68
2.2.3. Perfil de expresión de una cartera.....	70
CAPÍTULO III. METODOLOGÍA.....	78
3.1. BASES DE DATOS Y JUSTIFICACIÓN DE LAS MUESTRAS.....	78
3.2. ANÁLISIS PRELIMINAR DE LAS MUESTRAS .....	84
3.3. HIPÓTESIS DEL ESTUDIO .....	89
3.4. TÉCNICAS EMPLEADAS .....	92

3.3. DESCRIPCIÓN, ANÁLISIS Y RESULTADOS .....	98
3.3.1. Descripción.....	98
3.3.2. Análisis .....	114
3.3.3. Resultados .....	137
CAPÍTULO IV. CONCLUSIONES, IMPLICACIONES Y RECOMENDACIONES.....	146
4.1. CONCLUSIONES.....	146
4.2. IMPLICACIONES .....	152
4.3. RECOMENDACIONES.....	155
BIBLIOGRAFÍA .....	156
ANEXO I. ADN, ARN Y EL CICLO CELULAR.....	168
ANEXO I. ADN, ARN Y EL CICLO CELULAR.....	169
ANEXO II. EL CÁNCER .....	193
ANEXO III. RELACIÓN DEL CONJUNTO SELECCIONADOS .....	225

# **CAPÍTULO I. INTRODUCCIÓN**

## **1. 1. INTRODUCCIÓN**

El control del cáncer es, actualmente, una de las prioridades en salud pública dada la mortalidad que produce, así como el coste sanitario, económico y humano que acarrearán tanto la enfermedad como su tratamiento.

La curación de un proceso neoplásico no siempre es posible, por lo que durante las últimas décadas los estudios oncológicos realizados buscaban una mejor comprensión de la naturaleza del cáncer que permitiese obtener para los pacientes supervivencias cada vez mayores. Estos estudios se realizaban con el fin de conseguir técnicas quirúrgicas más depuradas, fármacos más eficaces y con menos efectos adversos, etc.

Hasta la fecha se han descrito más de 300 tipos de procesos neoplásicos diferentes. Son cientos, tal vez miles los genes que están implicados en cualquier tipo de cáncer. A su vez, los genes se encuentran influidos por complejos procesos bioquímicos que rigen el destino de las células, lo que complica aún más su estudio. A raíz de la secuenciación del genoma humano<sup>7</sup> (Proyecto Genoma Humano<sup>8</sup>) y los avances tecnológicos se inició una prometedora empresa consistente en cuantificar de forma simultánea todos los genes de un

---

<sup>7</sup> El Genoma Humano es el número total de cromosomas de un organismo, en este caso del ser humano. Los cromosomas contienen aproximadamente 40.000 genes.

<sup>8</sup> Proyecto internacional de investigación científica con el objetivo fundamental de determinar la secuencia de pares de bases químicas que componen el ADN e identificar y cartografiar los aproximadamente 40.000 genes del genoma humano desde un punto de vista físico y funcional. Fue fundado en 1990 en el Departamento de Energía y los Institutos de la Salud de los Estados Unidos, bajo la dirección de James D. Watson y presentado en abril del 2003.

individuo con la ayuda de los microarrays<sup>9</sup>. Desde entonces se están sucediendo numerosos estudios en todo el mundo que utilizan esta nueva técnica. Los investigadores, sin embargo, se encontraron con un problema para el que todavía no han encontrado una solución adecuada: el enorme volumen de datos del que ahora disponen. Este inconveniente radica en poder cribar de forma adecuada la información relevante de la inmensa cantidad de datos que ahora poseen. Esta novedosa manera de concebir y abordar el estudio del cáncer (hasta la aparición de los microarrays se estudiaba la relación existente entre unos pocos genes y ahora se pretende identificar la relación de cientos, tal vez de miles de genes simultáneamente), implica la entrada en escena de otras ramas del conocimiento como son la computación y la estadística. Situación que nos coloca en el punto de partida de un nuevo e interesante reto: dotar a los equipos de investigación oncológicos de un novedoso carácter multidisciplinar para poder integrar en sus estudios materias tales como: la biología molecular, la estadística o la gestión de complejas bases de datos, entre otras.

Esta etapa de integración de diversas áreas del conocimiento en su propio seno, lo lleva realizando, no sin esfuerzos y controversias, la economía, en general, desde hace décadas. La globalización ha supuesto mercados mucho más internacionales, competitivos e interrelacionados entre sí. Esto, unido a la continua aparición de nuevos productos financieros, eleva la complejidad en la valoración, tanto de los mercados como de los activos. Disciplinas como la econometría, las matemáticas financieras o el análisis financiero son buenos ejemplos de ello.

---

<sup>9</sup> Chip de superficie sólida a la cual se unen fragmentos de ADN. Mediante estos chips podemos observar casi de forma simultánea la expresión de todos los genes del genoma de un organismo. Para más información véase el Anexo IV.



La reciente y novedosa tecnología que se está usando ahora mismo en el campo de la oncología molecular<sup>10</sup>, deja abierta las puertas a otras ramas del saber para que puedan aportar sus conocimientos y experiencia, en sus respectivas áreas del conocimiento, en la lucha contra el cáncer.

## 1.2. OBJETIVOS

El propósito de esta Tesis Doctoral es ofrecer una técnica adicional, en el campo de la oncología molecular, que seleccione genes en función de la variabilidad en sus perfiles de expresión<sup>11</sup> y realice una nueva clasificación tumoral en virtud de un nuevo parámetro: la beta.

Para abordar esta cuestión se utiliza uno de los parámetros empleados en la caracterización del riesgo bursátil: la beta de un activo financiero. Trasladaremos este concepto a la oncología molecular.

Con esta finalidad los capítulos en los que se desglosa este trabajo de investigación son cuatro (el primero de los cuales es esta introducción) cuyo contenido se repasa brevemente a continuación.

El **Capítulo II** está dedicado al estudio de las principales teorías modernas de valoración de activos y su posible aplicación al campo de la oncología molecular. En primer lugar se resumen tres de ellas: Markowitz, CAPM y Modelo de Mercado. A continuación se presentan las técnicas que se van a emplear y la elaboración de la base de datos objeto de estudio.

---

<sup>10</sup> Estudio del cáncer a nivel molecular.

<sup>11</sup> Valor de la expresión de un gen que nos facilita el microarray.

En el **Capítulo III** se realiza el análisis de las 96 muestras (89 de pacientes con cáncer de colon y 7 muestras sanas) y se presentan los resultados obtenidos.

Finalmente, en el **Capítulo IV** se valoran los resultados obtenidos y se plantean una serie de sugerencias.

### 1.3. JUSTIFICACIÓN

La principal causa de muerte en los países desarrollados son las enfermedades cardiovasculares y oncológicas. Desde 1999 el cáncer ha superado a las cardiopatías como causa de muerte en personas menores de 85 años de edad<sup>12</sup>. El cáncer a día de hoy es un tema de trascendencia social de gran impacto. La probabilidad de presentar un cáncer durante toda la vida es de uno de cada dos para los varones y uno de cada tres para las mujeres, de acuerdo con la base de datos de Surveillance, Epidemiology and End-Results (SEER) de 1998-2000<sup>13</sup>.

Se calcula que en este siglo 1 de cada 3 españoles padecerán un cáncer, y aunque más del 50% podrá conseguir la curación, a la luz de los conocimientos actuales, casi el otro 50% morirá.

---

<sup>12</sup> Jemal A, Siegel R, Ward E, et al: Cancer statistics, 2006. CA Cancer J clin 2006; 56: 106-130.

<sup>13</sup> Gloeckler LA, Reichman ME, Riedel Lewis D, et al: Cancer survival and incidence from Surveillance, Epidemiology, and End results (SEER) program. Oncologist 2003; 8: 541-552.

	Hombre	Mujer	Ambos Sexos
1º	Próstata	Mama	Colorrectal
2º	Pulmón	Colorrectal	Próstata
3º	Colorrectal	Cuerpo de Útero	Pulmón
4º	Vejiga	Pulmón	Mama
5º	Estómago	Ovario	Vejiga

Tabla 1.1.

Fuente: Globocan 2012 (2014)

Elaboración: FCAECC.

El cáncer de colon es el tumor maligno de mayor incidencia en España. Cada año se diagnostican más de 25.000 casos nuevos y es responsable de cerca de 14.000 fallecimientos anuales.

En España, el cáncer colorrectal es el tercero más común en hombres con una tasa de incidencia de 39,7 casos y una mortalidad de 18 casos cada 100.000 habitantes. En mujeres, en cambio, ocupa la segunda posición con una incidencia de 23 casos y una mortalidad de 9,5 casos por 100.000 habitantes.

La incidencia del cáncer colorrectal (neoplasia objeto de estudio en este trabajo), el tumor más frecuente considerando la globalidad de hombres y mujeres (25.600 nuevos casos al año en España en 2014<sup>14</sup>), está en aumento, con un incremento anual de hasta el 2,6%, y constituye la primera causa de muerte oncológica cuando se analizan conjuntamente los casos en ambos sexos (14.000 fallecimientos anuales).

<sup>14</sup> Datos ofrecidos por el SEOM (Sociedad Española de Oncología Médica)

Y la realidad es que aún hoy, en el momento del diagnóstico, casi el 20% de los pacientes con cáncer colorrectal presenta metástasis<sup>15</sup> en el tumor, lo que significa que la enfermedad se ha extendido más allá del intestino grueso. En general, el 50% de los pacientes con esta enfermedad desarrolla metástasis y finalmente fallece. Sin embargo, su incidencia y mortalidad son inferiores a la media de los países europeos, especialmente en comparación con los países del Norte de Europa. Además, los resultados que actualmente pueden conseguirse en este tumor, gracias a su abordaje multidisciplinario, han hecho que haya cambiado su historia natural.

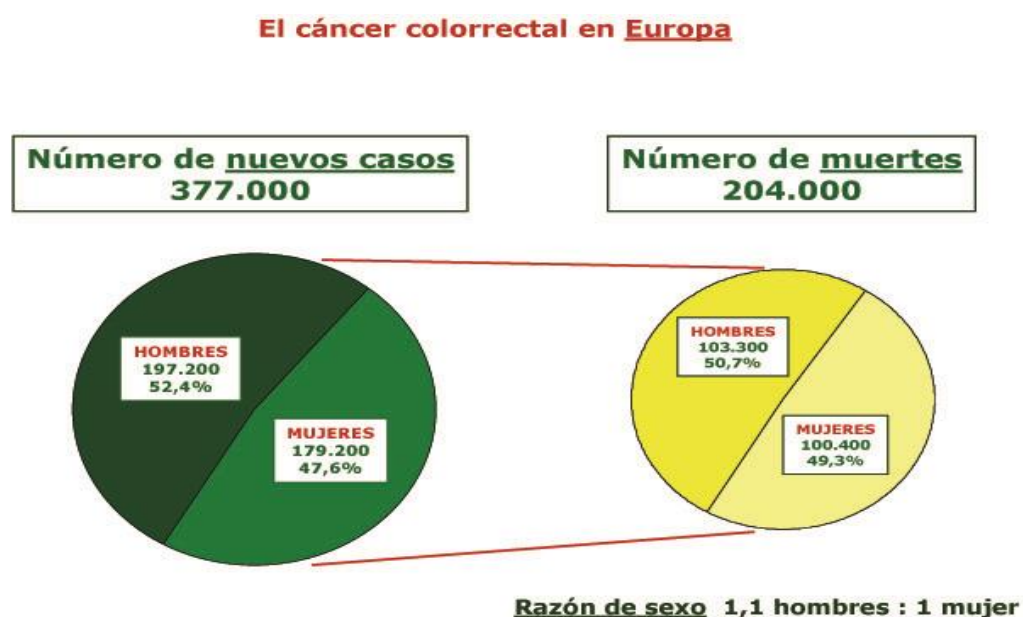


Imagen 1.2.

Fuente: Globocan 2002 (2005)

Elaboración: FCAECC.

<sup>15</sup> Propagación de un foco canceroso a un órgano o tejido distinto de aquel en el que se inició. Para más información véase el Anexo II.

### **El cáncer colorrectal en el mundo**

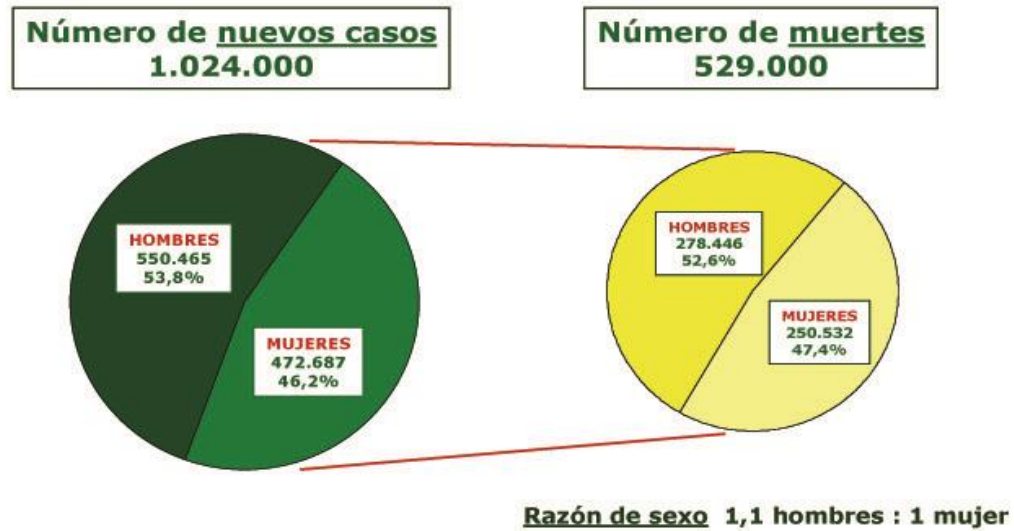


Imagen 1.3.

Fuente: Globocan 2002 (2005)

Elaboración: FCAECC.

Basta mencionar estos datos aportados por la SEOM<sup>16</sup> (Sociedad Española de Oncología Médica) para poner de manifiesto la trascendencia social que tiene la investigación en este campo y que no deben escatimarse esfuerzos en la lucha contra esta enfermedad.

A raíz de la secuenciación del genoma humano (14 de abril del 2003) la biomedicina se vio desbordada por la inmensa cantidad de información que este hito trajo consigo. Se hizo preciso pues, la colaboración de otras áreas del conocimiento. Algunas disciplinas, como la informática o la estadística, han cobrado un papel capital en la investigación oncológica. Fruto de esto es la Bioinformática, una disciplina científica emergente que utiliza las tecnologías de

<sup>16</sup> Datos hechos públicos en el XI Congreso Sociedad Española de Oncología Médica celebrado los días 3-5 de octubre del 2007 en el IFEMA, Madrid.

la información para organizar, analizar y distribuir información biológica con la finalidad de responder a preguntas complejas en biología. Bioinformática es un área de investigación multidisciplinaria, la cual puede ser ampliamente definida como la interfase entre dos ciencias, Biología y Computación, y está impulsada por la incógnita del genoma humano y la promesa de una nueva era en la cual la investigación genómica puede ayudar de forma espectacular a mejorar la condición y calidad de vida humana.

El descubrimiento de que no son un único gen o genes independientes entre sí, sino la sinergia conjunta de cientos, tal vez de miles de ellos los responsables de enfermedades como el cáncer ha supuesto un cambio fundamental en el diseño de nuevas estrategias para comprender mejor sus mecanismos de funcionamiento. Esta nueva concepción de la genética llamada “genómica” trata, entre otros fines, de determinar relaciones entre genes y cuantificar la expresión simultánea de miles de ellos.

Los genes que controlan la división y proliferación celular son responsables del cáncer. El cáncer es un fenómeno altamente ligado al reloj biológico celular y por tanto al envejecimiento; por lo que, al aumentar la esperanza de vida de la población occidental, aumentará a su vez la incidencia de procesos neoplásicos malignos.

Cuando la difícil empresa de identificar genes clave se consigue, suele desembocar en el desarrollo de nuevos fármacos que actúen sobre esos genes.

Uno de los aspectos capitales de la investigación oncológica es conocer los mecanismos y genes que regulan el ciclo de vida celular de las células malignas y las proporcionan determinadas características clave para el desarrollo de la enfermedad, como pueden ser: un crecimiento mucho más rápido que las células sanas, mayor supervivencia en medios hostiles, capacidad para hacer metástasis, capacidad para eludir tantos los mecanismos de defensa del cuerpo como para evitar la apoptosis (muerte programada de la célula), o la capacidad que en muchos casos presentan las células cancerosas para ofrecer resistencia a los tratamientos farmacológicos. Cuando una de estas claves es descubierta en alguno de los distintos tipos de cáncer los resultados suelen ser prometedores. Un ejemplo es el cáncer de mama. Aunque el cáncer de mama es el tumor más frecuente en la mujer y la primera causa de muerte por enfermedad oncológica, los avances médicos, que se han ido sucediendo sin interrupción en los últimos años, están consiguiendo aumentar el número de curaciones, es decir, la supervivencia y calidad de vida de las pacientes. Según datos de la Asociación Española contra el Cáncer, cada año se diagnostican en nuestro país 16.000 casos nuevos de cáncer de mama, lo que representa un 24,6% del total de tumores detectados en mujeres españolas. No obstante, los expertos se muestran optimistas porque, según los estudios del Ministerio de Sanidad, se aprecia una importante disminución de la mortalidad en los últimos diez años, a un ritmo del 1,4% anual. En la actualidad, la supervivencia del cáncer de colon a los cinco años se eleva a un 53%, para los hombres y un 54,7% para las mujeres según el estudio EUROCORE-3<sup>17</sup> (ver tabla 1.4). Qué duda cabe que cualquier línea de investigación que pueda aportar un rayo de luz sobre tan complicado tema debe ser estudiado.

---

<sup>17</sup> European Cancer Registry Study of Survivals and Care of Cancer patients.

Como consecuencia del conocimiento cada vez más preciso que se tiene sobre los mecanismos moleculares que rigen los procesos celulares que originan el nacimiento y desarrollo del cáncer, los grupos de investigación han ido cobrando un cierto carácter multidisciplinar. Es habitual, hoy en día, que no sólo los médicos, biólogos, químicos o farmacéuticos integren estos equipos sino también matemáticos, estadísticos e informáticos. A medida que los avances se suceden son mayores las necesidades de conocimientos que de otras áreas se necesitan. Este cambio sustancial que está experimentando la investigación biomédica lo tuvo el análisis financiero hace décadas.



Cáncer	HOMBRES		MUJERES	
	ESPAÑA	EUROPA	ESPAÑA	EUROPA
Labio	99,2	92,5	94,9	90,3
Lengua	37,5	34,9	56,1	52,2
Gl. Salivares	58,8	51	87,7	68,7
Cavidad Oral	45	40,6	51	57
Orofaringe	29	28,7	44	43,5
Nasofaringe	48,8	40	52	49,3
Hipofaringe	24,5	23,2	48,6	32,3
Esófago	12,6	8,5	23,2	10,5
Estómago	26,8	20	30,4	25,4
Colorrectal	53	47,6	54,7	50,5
Hígado	10,4	6,2	11,6	6,7
Vías Biliares	17,3	12,3	19,6	11,6
Páncreas	5,1	3,8	6,3	4,6
Cavidad Nasal	41	43,8	63,2	48
Laringe	66,7	60,7	80	59,4
Pulmón	12,4	9,7	12,8	9,6
Pleura	13	4,4	20	5,5
Hueso	54	53	64,8	56,3
Tejidos Blandos	46,8	54,2	59,3	54,2
Melanoma	74	74,8	89,8	84,3
Mama	82,8	75,5	78	76,1
Cuello de Útero	-	-	68,7	62,1
Cuerpo de Útero	-	-	75,4	76
Ovario	-	-	43,2	36,7
Vulva y Vagina	-	-	59,5	52,2
Próstata	65,5	65,4	-	-
Testículo	87,4	91,4	-	-
Pene	78,7	68,9	-	-
Vejiga	75,4	69,5	70,4	67,1
Riñón	56,1	54,2	60,7	57,2
Encéfalo	18,8	16,4	17,6	18,5
Tiroides	82	71,8	85,7	81,4
Linfomas NH	53	47,7	55	53,7
Linfomas Hodgkin	72,8	75,2	76,1	81,5
Mieloma Múltiple	34	28,5	37,5	33
LLA	31,8	24,2	48	21,6
LLC	81,7	62,2	75	66,4
LMA	13,6	12,7	15,7	13,4
LMC	35,3	30,5	48	37
Todas las Leucemias	46,6	36,5	44,1	37,2
<b>TODOS (excepto piel)</b>	<b>44</b>	<b>39,8</b>	<b>56,4</b>	<b>51,2</b>

Cuadro 1.4. % de supervivencia a los 5 años. Casos diagnosticados entre 1.990 y 1.994, con seguimiento hasta finales de 1.999. Comparación entre España y la media europea.

Fuente: Enrocare

Elaboración: FCAECC.

Cáncer	Incidencia <sup>1</sup>			Mortalidad <sup>1</sup>			Prevalencia a 5 años <sup>2</sup>		
	Número	(%)	ASR (W)	Número	(%)	ASR (W)	Número	(%)	ASR (W)
Labio y cavidad oral	4098	1.9	4.7	1117	1.1	1.2	11811	2.0	29.8
Nasofaringe	350	0.2	0.5	188	0.2	0.2	909	0.2	2.3
Otros faríngeos	1530	0.7	2.1	765	0.7	1.0	3792	0.7	9.6
Esófago	2090	1.0	2.5	1728	1.7	1.9	2238	0.4	5.7
Estómago	7810	3.6	7.8	5389	5.2	4.9	12611	2.2	31.8
Colorrectal	32240	15.0	33.1	14700	14.3	12.3	89705	15.4	226.3
Hígado	5522	2.6	5.9	4536	4.4	4.3	5506	0.9	13.8
Vesícula biliar	2002	0.9	1.7	1174	1.1	0.9	1905	0.3	4.8
Páncreas	6367	3.0	6.3	5720	5.6	5.5	3437	0.6	8.7
Laringe	3182	1.5	4.1	1321	1.3	1.5	11200	1.9	28.3
Pulmón	26715	12.4	30.3	21118	20.6	22.8	28148	4.8	71.0
Melanoma	5004	2.3	6.9	967	0.9	1.0	19792	3.4	49.9
Sarcoma de Kaposi	316	0.1	0.4	24	0.0	0.0	880	0.2	2.2
Mama	25215	11.7	67.3	6075	5.9	11.9	104210	17.9	516.2
Cuello de útero (cérvix)	2511	1.2	7.8	848	0.8	2.1	8867	1.5	43.9
Cuerpo del útero	5121	2.4	11.6	1211	1.2	1.9	19272	3.3	95.5
Ovario	3236	1.5	7.7	1878	1.8	3.7	7925	1.4	39.3
Próstata	27853	12.9	65.2	5481	5.3	8.6	102559	17.6	527.3
Testículo	823	0.4	3.5	42	0.0	0.1	3242	0.6	16.7
Riñón	6474	3.0	7.8	2295	2.2	2.2	18693	3.2	47.2
Vejiga	13789	6.4	13.9	5007	4.9	4.0	47225	8.1	119.2
Cerebro, sistema nervioso	3717	1.7	5.1	2668	2.6	3.3	3187	0.5	8.0
Tiroides	2059	1.0	3.4	286	0.3	0.3	8097	1.4	20.4
Linfoma Hodgkin	1150	0.5	2.3	212	0.2	0.3	3644	0.6	9.2
Linfoma no Hodgkin	6130	2.8	7.5	2337	2.3	2.2	16342	2.8	41.2
Mieloma múltiple	2420	1.1	2.3	1675	1.6	1.4	5730	1.0	14.5
Leucemia	5190	2.4	6.5	3212	3.1	3.0	11040	1.9	27.9
Todos los cánceres excl. cáncer de piel no melanoma	215534	100.0	249.1	102762	100.0	98.1	581688	100.0	1467.9

Cuadro 1.5. Estimación sobre incidencia, mortalidad y prevalencia a 5 años en ambos sexos de los tumores más frecuentes.

Fuente: GLOBOCAN 2012.

Elaboración: FCAECC.

Dos de los campos objeto de estudio en finanzas, la valoración de activos financieros y el análisis bursátil, han experimentado un gran auge durante la segunda mitad del siglo XX.

Existen dos grandes áreas en las que tradicionalmente se han dividido las finanzas: la valoración de activos y la denominada “corporate”, dedicada a la

estructura financiera de la empresa. Dentro de la primera podrían establecerse varias áreas de conocimiento para la inversión en Bolsa, como por ejemplo: el análisis técnico, el fundamental, las denominadas teorías modernas y la teoría de opciones.

El mercado bursátil ha sido objeto de una gran atención por parte de los académicos y del público en general durante las últimas décadas. El papel de la bolsa dentro de la economía de un país cada vez está cobrando más relevancia, por las posibilidades de financiación que supone para la empresa, por la importancia del ahorro dentro de las economías domésticas, y por ser el termómetro del estado de la economía

Hasta los años cincuenta, el análisis de las inversiones bursátiles, con el objetivo de seleccionar los valores idóneos para invertir en ellos, se realizaba tomando como variable de decisión la rentabilidad que el título y/o la cartera habían producido en periodos anteriores. A principios del siglo XX, el análisis técnico era una herramienta ampliamente empleada en el estudio de valores, y el análisis fundamental había cobrado gran importancia a raíz de la Gran Depresión del 29. Precisamente esa crisis hizo notar la necesidad de tomar decisiones de inversión en bolsa con criterios no especulativos, sino basados en el estudio de los estados financieros de las empresas y del sector en el que desarrollan su actividad.

Lo importante no es conocer la rentabilidad que hemos obtenido con una inversión, sino la rentabilidad que podemos obtener realizándola. Estamos

hablando del futuro, y el futuro es incierto. La introducción de esa “incertidumbre” en el análisis bursátil, fue lo que propuso en 1952 Harry Markowitz en su artículo “Portfolio Selection”<sup>18</sup>. Aunque el nacimiento de esta teoría se sitúa en 1952, no es hasta 1959 cuando el propio Markowitz publicó su artículo<sup>19</sup>.

La rentabilidad es la ganancia neta por unidad monetaria invertida, ganancia que puede derivarse del cobro de dividendos (o intereses en el caso de renta fija), ventajas de las ampliaciones de capital y plusvalías por aumento de la cotización de los títulos. En cuanto al riesgo, al ser la rentabilidad una variable aleatoria, la idea de Markowitz fue medirlo mediante la varianza de la rentabilidad o desviación media cuadrática de esta variable respecto a su media. Diseñó un marco teórico, denominado teoría de carteras, para analizar el problema frente al cual se encuentra todo inversor en bolsa: invertir unos recursos limitados en una combinación de valores que formen una cartera “eficiente”.

La denominada teoría de carteras cuenta entre sus méritos el haber introducido los métodos cuantitativos y la perspectiva científica en el estudio de las inversiones. En este sentido, uno de sus principales frutos ha sido el conocido Modelo de Valoración de Activos Financieros o CAPM (Capital Asset Pricing Model), que se ha convertido en herramienta fundamental para las finanzas, tanto en su ámbito profesional como académico.

---

<sup>18</sup> Markowitz H., Portfolio Selection, “Journal of Finance”, vol .VII, num. 1, marzo 1952, págs. 77-91.

<sup>19</sup> Markowitz H., Portfolio Selection: Efficient Diversification of Investment, John Wiley, Nueva York 1959, pág 38.

El modelo CAPM, desarrollado a mediados de los sesenta por Sharpe, Lintner y Mossin, es de carácter predictivo y expresa analíticamente la razonable idea de que un inversor adverso al riesgo sólo adquirirá activos arriesgados si se ve adecuadamente compensado por la asunción de dicho riesgo, y que su rentabilidad será mayor cuanto mayor sea el riesgo del activo de que se trate – medido por la beta -.

Si hacemos un análisis retrospectivo en el tiempo, desde los años 70 hasta nuestros días, es “volatilidad” la palabra que mejor define el comportamiento de los mercados financieros, nacionales e internacionales, durante las tres últimas décadas. Este incremento de la volatilidad en los mercados financieros tiene su razón de ser en los distintos acontecimientos que se han sucedido en el contexto económico mundial. Y es que, la globalización de la economía, la integración de los mercados y los avances tecnológicos, a los que también hemos asistido de forma paralela, han contribuido a propagar sobremanera los efectos de tales acontecimientos. Esta situación de alta volatilidad ha potenciado, en gran medida, el desarrollo de la ingeniería o innovación financiera.

Beta, también llamado coeficiente de volatilidad, es componente fundamental del riesgo sistemático, parte integrante a su vez del riesgo total de un activo, el cual viene dado por la varianza de su rentabilidad, y es susceptible de ser expresado, por tanto, como suma del riesgo sistemático (parte de la variabilidad de la rentabilidad del activo dependiente de la evolución del mercado, en este caso del tumor que siempre está presente porque difícilmente se puede

evitar que la rentabilidad de los activos se vea influida por los movimientos del mercado en el que se negocian), y el riesgo específico o no sistemático, depende exclusivamente de las características particulares de la empresa emisora del título y puede eliminarse o reducirse mediante diversificación).

La beta de un activo financiero se calcula como el cociente entre el riesgo del activo en cuestión en relación al de mercado (medido por las varianzas de las rentabilidades) y el riesgo de mercado (dado por la varianza de su rendimiento), y mide la contribución marginal del activo al riesgo de mercado.

Beta también es la pendiente de la recta que relaciona la rentabilidad de un activo con la cartera de mercado o, en su defecto, con la de un índice económico utilizado como aproximación. Recta que se conoce con el nombre de línea característica o modelo de mercado de Sharpe, y que es importante, además, porque a partir de ella se obtiene el desglose del riesgo total en sus componentes sistemático y no sistemático; y porque pone de relieve el hecho de que beta es un concepto con entidad propia, ya que surge con independencia del CAPM.

Las carteras de inversiones con acciones acostumbran a mantener un gran número de títulos, de forma que la elaboración diaria de la matriz de varianzas y covarianzas puede llegar a ser un trabajo sumamente complejo. Por esta razón, el cálculo del riesgo de las acciones suele realizarse mediante la beta de la cartera, lo que significa que capturamos el riesgo sistemático. La beta de una acción es la variación en el precio de esta acción ante variaciones de un

índice de mercado. Por lo tanto, lo que haremos será sintetizar el riesgo de una cartera de acciones en términos de riesgo de su índice.

La beta del mercado es igual a la unidad. Ello nos permite hacer una primera clasificación de activos financieros de cara a la gestión del riesgo: activos de alto riesgo específico, con beta mayor que la unidad; y activos con bajo riesgo específico, con beta menor que la unidad; activos anticíclicos, con beta negativa. El rendimiento esperado de estos tipos de activos dependerá linealmente del riesgo específico de cada activo, medido por las betas, de manera que mayor a beta, mayor riesgo respecto al propio mercado, y mayor rendimiento operado. Una política de inversión agresiva supondrá la selección de activos con betas mayores que uno; por tanto, ante subidas en el mercado los activos se comportarán mejor que la media, mientras que en caídas del mercado se debe esperar, asimismo, peores resultados. Una política conservadora, basada en la beta de los activos, nos aconsejará la selección de activos con beta menor que la unidad, que perderá parte de las subidas, pero también suavizará la pendiente de las bajadas. La selección de valores acíclicos, aunque con rentabilidades medias esperadas por debajo del tipo de interés libre de riesgo, permite resguardarse de las caídas de los mercados en espera de mejores condiciones para situarse en valores que sigan la senda de la economía.

El coeficiente beta cuenta con una propiedad muy importante, de cara a la utilización del mismo como herramienta de gestión, como es la aditividad. La beta de una cartera es la combinación lineal de las betas de los activos individuales

que la componen. Así, una determinada cartera  $C_p$  está compuesta por  $n$  activos con las correspondientes ponderaciones recogidas en el vector  $A_p$ .

$$A_p = (a_1, a_2, \dots, a_j, \dots, a_n)$$

Se puede obtener la beta de la cartera de acuerdo con la siguiente expresión:

$$\beta_p = a_1\beta_1 + a_2\beta_2 + \dots + a_j\beta_j + \dots + a_n\beta_n$$

$$\beta_p = \sum_{i=1}^n a_i\beta_i$$

Ello permite que podamos hablar de carteras agresivas, si su beta es mayor que la unidad; carteras conservadoras, si su beta es menor que uno; y de carteras réplicas del mercado, si su beta es igual a la unidad.

La sencillez del modelo ha facilitado su empleo generalizado en diversos aspectos de gestión financiera. Su aplicación más inmediata, hace honor a su nombre, es la valoración de activos, concretamente en un horizonte temporal de un período. También se utiliza con profusión en el cálculo del coste de capital de una empresa, a efectos de selección de inversiones, y en la evaluación de carteras, o medida de la “performance” de las mismas.



#### **1.4. LIMITACIONES**

Uno de los principales inconvenientes con los que se encuentran los investigadores en el campo de la oncología molecular son las muestras objeto de estudio.

- Los pacientes presentan una gran heterogeneidad: diferentes edades, distintas patologías, distintos estadios del tumor, etc.
- Las muestras de los pacientes son tomadas en el momento de la intervención quirúrgica sin posibilidad de disponer de muestras posteriores que nos indicaran la evolución de los genes objeto del análisis.
- No es posible, en muchos casos, contar con la evolución clínica de los pacientes, su respuesta a los tratamientos suministrados o su índice de supervivencia.

Además, el número de muestras analizadas no suelen ser elevadas.

## **CAPÍTULO II. FUNDAMENTOS CONCEPTUALES**

### **2.1. REVISIÓN DE LOS PRINCIPALES MODELOS DE LA TEORÍA DE CARTERAS**

El autor de la presente Tesis no considera necesario, pues sería propósito de otra Tesis Doctoral, realizar un análisis minucioso y pormenorizado de las principales teorías de carteras, por lo que se limitará a una visión genérica de alguna de ellas (Markowitz, CAPM y Modelo de Mercado).

#### **2.1.1. Markowitz**

La teoría de carteras fue desarrollada por Harry Markowitz<sup>20</sup>, su principal aplicación es la creación de un modelo en el que se recogen los rasgos fundamentales de lo que teóricamente debería ser el comportamiento de un inversor. En 1959 el propio Markowitz desarrolló su formulación inicial en su nueva obra<sup>21</sup>, en un modelo más detallado.

Dado un conjunto de activos financieros, Markowitz plantea cuáles son las mejores combinaciones entre dichos activos y, de entre ellas, cuál escoger a la hora de tomar la decisión de invertir. Para tratar de dar respuesta a esta cuestión propone un modelo matemático que pretende maximizar la utilidad esperada de un inversor racional y adverso al riesgo resolviendo el problema de cómo repartir su presupuesto de inversión entre diferentes activos financieros arriesgados que se negocian en el mercado. Este modelo permite encontrar mejor la cartera para

---

<sup>20</sup> Markowitz H., Portfolio Selection, "Journal of Finance", vol .VII, num. 1, marzo 1952, págs. 77-91.

<sup>21</sup> Markowitz H., Portfolio Selection: Efficient Diversificacion of Investment, Jhon Wiley, Nueva York 1959, pág 38.

ese inversor, es decir, la composición de la cartera que maximiza su utilidad esperada, función que depende del rendimiento esperado y de la varianza de la rentabilidad de la cartera.

Como principal aportación del modelo cabe destacar que recoge de forma explícita el comportamiento racional del inversor con base al criterio de decisión media-varianza, que consiste en buscar aquella composición de la cartera que maximice su rendimiento para un determinado nivel de riesgo o, alternativamente, que minimice su riesgo para un rendimiento dado.

El enfoque teórico puede utilizarse de dos maneras. Un modelo normativo, es una guía de acción: indica de qué modo se deben tomar las decisiones, ayuda al decisor a llevar a cabo sus objetivos. Un modelo positivo sirve para predecir el entorno; describe la forma en la que se toman las decisiones y las relaciones entre los conceptos tales como: precios, cantidades, etc. La teoría de carteras está dentro del enfoque normativo, describe la forma en la que debería comportarse un inversor teórico y tal vez un gestor de inversiones, en ambiente de incertidumbre.

Se podría plantar el problema de la selección de carteras es tres etapas independientes:

Análisis de los títulos: para esta primera etapa es necesario predecir el comportamiento futuro de los títulos y, como sabemos, el futuro es incierto lo que nos lleva a asumir un riesgo.

Análisis de las carteras: para ello es necesario tener en consideración en análisis realizado en la etapa anterior.

Selección de carteras: se elegirán de entre todas las posibles carteras, aquellas que satisfagan, en un aspecto u otro, al inversor.

Herrero de Egaña (1999, pp. 87-88) sostiene que Markowitz asocia riesgo y rendimiento e introduce conceptos como rendimientos esperados y medidas de dispersión en la distribución de los mismos, así como la covarianza entre los títulos. Piensa que el inversor considera los rendimientos futuros como algo deseable y la variación de los mismos algo que le produce insatisfacción. La forma de reducir las variaciones en el rendimiento de los valores, el riesgo, es la diversificación. Los rendimientos de un título o de una cartera se describen por medio de una variable aleatoria con una distribución de probabilidad para el período de referencia conocida por el inversor. El valor medio o esperanza matemática de dicha variable se acepta como medida de rentabilidad de la inversión. La medida del riesgo es la varianza o la desviación estándar de la variable aleatoria en la que nos basamos para obtener rendimiento. La conducta del inversor, al que se le supone adverso al riesgo, se basa en el binomio rendimiento y riesgo, prefiriendo aquellas inversiones con mayor rendimiento y menor riesgo. Existen dos conclusiones fundamentales que se pueden extraer de todo lo mencionado hasta el momento: la primera que la diversificación es una consecuencia natural de la regla que rige el comportamiento del inversor, porque si no tuviera en cuenta el riesgo invertiría solamente en aquellas acciones que le proporcionaran mayores rendimientos; la segunda es que diversificará entre las combinaciones que le proporcionen mayores rendimientos<sup>22</sup>, la ley de los grandes números aseguraría que el rendimiento de la cartera fuera igual al rendimiento esperado.

---

<sup>22</sup> Markowitz H., Portfolio Selection, "Journal of Finance", vol .VII, num. 1, marzo 1952, pág 79.

Una cartera es eficiente en el sentido Markowitz cuando se cumplen simultáneamente las dos condiciones siguientes<sup>23</sup>:

Para su nivel de rendimiento esperado, no existe otra cartera que tenga un riesgo menor.

Para el riesgo que soporta, no existe otra cartera que ofrezca un rendimiento mayor esperado<sup>24</sup>.

Un inversor, racional y averso al riesgo, únicamente se planteará la selección de carteras eficientes. Aunque este sería el siguiente paso a estudiar (dado que no todas las carteras serán de utilidad para el inversor es necesario articular algún procedimiento de búsqueda de las carteras que nos proporcionen mayores rendimientos con menor varianza, o los de menor varianza para una rentabilidad dada); vamos a detenernos en las razones que inducen a elegir los dos conceptos mencionados como representativos del rendimiento y el riesgo.

Tal y como señala Fernando Bolado (2006, pp 54-63) en un universo de dos parámetros  $(E, \sigma^2)$  y dos instantes de tiempo (hoy y un momento futuro), un inversor estará interesado en colocar sus ahorros, o parte de ellos, comprando hoy acciones de la empresa  $i$  por un precio unitario cierto  $(P_{i0})$ ,

---

<sup>23</sup> La diferencia entre el plano media-desviación típica y el plano media-varianza está en la escala de abscisas. Así pues, para ver si una cartera es eficiente se pueden usar indistintamente ambos criterios. El resultado debe ser el mismo, ya que una cartera eficiente lo es en cualquiera de los dos planos.

<sup>24</sup> Markowitz exige que la cartera sea legítima para calificarla como eficiente. Se entenderá que una cartera es legítima cuando las proporciones del presupuesto invertido en los distintos títulos que la componen sean positivas, esto es, que no se produzcan ventas en descubierto de los títulos.

si tiene la esperanza de que suba su cotización y las pueda vender dentro de un período de tiempo por un precio unitario aleatorio ( $\tilde{P}_{it}$ ) más elevado. Si, además, a lo largo de dicho período la empresa va a repartir dividendos por acción inciertos ( $\tilde{d}_{it}$ ), la rentabilidad aleatoria<sup>25</sup> que finalmente obtendrá el accionista en el horizonte de planificación  $t$  que se ha fijado será<sup>26</sup>:

$$\tilde{R} = \frac{(\tilde{P}_{it} - P_{i0}) + \tilde{d}_{it}}{P_{i0}}$$

El rendimiento o rentabilidad de un título ( $\tilde{R}$ ), expresada en tanto por uno o tanto por ciento, será igual al cociente entre la cantidad de euros que percibe el accionista vía plusvalías ( $(\tilde{P}_{it} - P_{i0})$ ), o vía cualquier otro flujo neto de caja ( $\tilde{d}_{it}$ ), y el importe de su inversión inicial ( $P_{i0}$ ).

Como cualquier variable aleatoria podemos calcular diferentes estadísticos de la distribución de los rendimientos de un título. La media o esperanza matemática se obtiene sumando todos los posibles valores que puede tomar la variable del título  $i$  previamente ponderados por sus respectivas probabilidades de ocurrencia. En el caso discreto toma la siguiente expresión:

---

<sup>25</sup> Para distinguir una magnitud cierta de una magnitud aleatoria, hemos añadido una tilde encima de las variables que tiene un comportamiento probabilístico.

<sup>26</sup> En la práctica, la variable aleatoria  $\tilde{d}_{it}$  debe recoger no sólo los dividendos por acción que el accionista en su horizonte de planificación, sino cualquier otro cobro proveniente de reducciones del nominal, amortizaciones de acciones, etc. Igualmente, si se produce una ampliación de capital, el desembolso necesario para acudir a ella deberá ser restado. En definitiva,  $\tilde{d}_{it}$  es el flujo neto de caja por acción que percibirá el inversor en el próximo período por cualquier concepto distinto a la plusvalía o minusvalía que se pueda producir al comparar el precio de venta del activo financiero  $i$  con su precio de compra.

$$E(\tilde{R}_i) = \sum_{j=1}^m R_{ij} P_j$$

donde:

$R_{ij}$  es la rentabilidad del activo i en el estado de la naturaleza j

$$(i=1,2, \dots, N \quad \text{y} \quad j=1,2, \dots, m)$$

$P_j$  es la probabilidad de que ocurra el estado de la naturaleza j.

La varianza mide la variabilidad de la rentabilidad del título i. Cuanto mayor sea ésta, más dispersos estarán los posibles valores de la rentabilidad, siendo más incierta y arriesgada la rentabilidad del activo financiero de que se trate:

$$\sigma^2(\tilde{R}_i) = E\left[\left(\tilde{R}_i - E(\tilde{R}_i)\right)^2\right] = E(\tilde{R}_i^2) - \left[E(\tilde{R}_i)\right]^2 = \sum_{j=1}^m \left(R_{ij} - E(\tilde{R}_i)\right)^2 P_j$$

Para obtener la varianza, en primer lugar, se calculan las diferencias entre cada posible valor de la rentabilidad y la rentabilidad media; posteriormente, esas desviaciones se elevan al cuadrado para que no se compensen entre sí<sup>27</sup> y, finalmente, se multiplican esos resultados por sus respectivas probabilidades de ocurrencia con la finalidad de obtener un promedio.

Si, como es lo más habitual, la rentabilidad media ( $E(\tilde{R}_i)$ ) está expresada en términos de porcentajes o de tanto por uno, entonces  $\sigma^2(\tilde{R}_i)$ , al tratarse de

---

<sup>27</sup> Dado que habrá diferencias positivas (+) y negativas (--), si no se elevan al cuadrado las desviaciones, al obtener el promedio se podrían anular entre sí, dando una falsa impresión de que la variabilidad de los rendimientos es nula, cuando en realidad, sí que hay variabilidad y ésta puede ser muy alta.

desviaciones al cuadrado, vendrá expresada como un porcentaje o como un tanto por uno al cuadrado. Por este motivo se suele trabajar con la desviación típica de los rendimientos como medida de dispersión o de riesgo consiguiendo así que tanto la media como la desviación típica estén expresadas en porcentajes. La desviación típica de una variable aleatoria es la raíz cuadrada positiva de la varianza.

$$\sigma(\tilde{R}_i) = \sigma_i = \sqrt{\sigma^2(\tilde{R}_i)}$$

La varianza de los rendimientos de una cartera  $p$  también puede expresarse de forma matricial como sigue:

$$\sigma^2(\tilde{R}_p) = (w_1, w_2, \dots, w_N) \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2N} \\ \dots & \dots & \dots & \dots \\ \sigma_{1N} & \sigma_{2N} & \dots & \sigma_N^2 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_N \end{pmatrix} = x' V x$$

La búsqueda de la cartera se lleva a cabo, Lassala Navarré (2006, pp 79-106), en cuatro etapas.

- A. Delimitación del conjunto de posibilidades de inversión
- B. Determinar las carteras eficientes, entendiendo por tales, las que proporcionan la máxima ganancia para un riesgo determinado o las que para un mínimo riesgo nos dan una ganancia dada.



C. Especificar las curvas de indiferencia entre ganancia y riesgo, cuya forma dependerá de la función de utilidad, que será distinta para cada inversor. Markowitz supuso que los inversores tenían aversión por el riesgo.

D. La determinación de la cartera óptima, se hace hallando un punto en el cual la curva de carteras eficientes es tangente a una de las curvas de indiferencia. Cualquier otro punto de las carteras eficientes se corresponde con una curva de indiferencia de menor índice de utilidad.

Veamos cada uno de estos pasos con más detenimiento.

**A) Delimitación del conjunto de posibilidades de inversión.** Dado que suponemos que los individuos escogen sus carteras exclusivamente de acuerdo con el rendimiento esperado y la desviación típica de sus inversiones, para poder obtener el conjunto de oportunidades de inversión con múltiples activos en el entorno media-varianza tendremos que caracterizar todas las posibles combinaciones de rendimiento esperado y desviación típica que pueden lograrse dado un número  $n$  de activos con riesgo.

Así pues en esta primera etapa se analiza el conjunto de activos que se negocian en el mercado. En concreto, se estiman para cada título su rentabilidad esperada, su riesgo (medido éste por la varianza o la desviación típica), y sus covarianzas con el resto de títulos que se pueden formar.

Una vez conocidas las esperanzas, la desviación típica de la rentabilidad de los  $n$  títulos que se negocian en el mercado y los pares de covarianzas entre rendimientos de cualquier par de títulos y, por tanto, del infinito número de carteras que se pueden formar con dichos títulos, podemos representar gráficamente todas esas posibles combinaciones de rendimiento esperado y riesgo de activos individuales y de carteras en los ejes de coordenadas rentabilidad esperada-desviación típica. De esta forma, como se muestra en la figura 2.1 obtendremos el conjunto de oportunidades de inversión que viene dado por el mercado tal y como aprecia cada inversor.

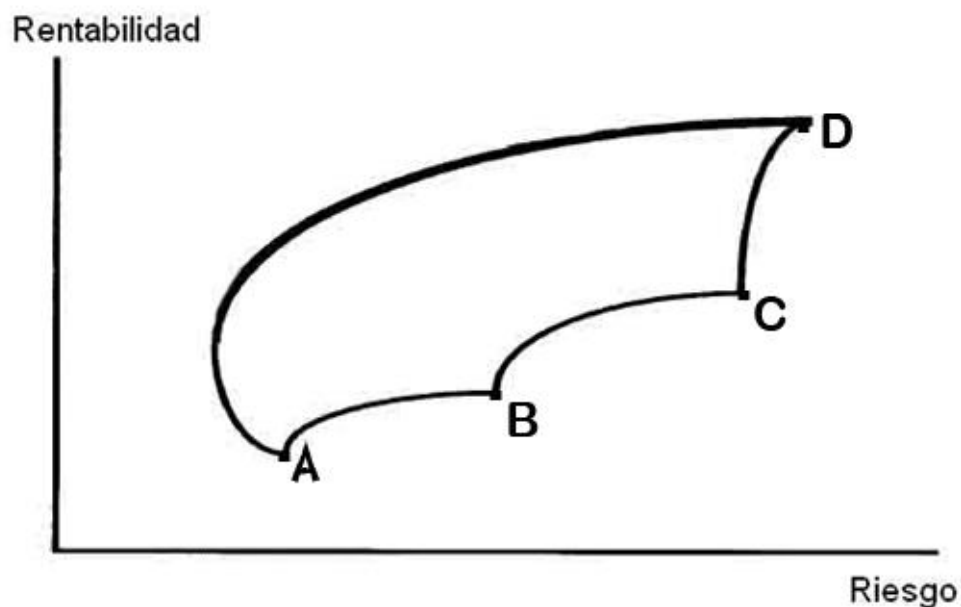


Figura 2.1. Conjunto de oportunidades de inversión con cuatro activos.  
Fuente: elaboración propia.

El conjunto posible representa todas las oportunidades de inversión en el mercado. Es un conjunto acotado, si bien en él existen infinitos puntos que evidencian infinitas posibilidades de inversión.

El conjunto de posibles carteras que se pueden formar con los distintos activos es enorme y ocupa numerosos puntos del espacio media-desviación típica. Una vez que hemos delimitado el conjunto de oportunidades de inversión, ahora el inversor deberá considerar cuál es la cartera, es decir, la combinación de inversiones arriesgadas de entre las posibles, que más le satisface y para ello tendrá que buscar aquella combinación rendimiento-riesgo que maximice su utilidad esperada.

**B) Determinación de la cartera eficiente.** Para generar la frontera eficiente podemos seguir dos caminos:

Fijar los niveles de riesgo, medidos por la desviación típica (o varianza) y determinar para cada uno de ellos la cartera que proporcione la máxima rentabilidad esperada.

Especificar los valores que puede alcanzar la rentabilidad esperada y determinar entonces para cada uno de esos valores la cartera de menor riesgo.

Una cartera será ineficiente si es posible obtener mayor rendimiento sin incrementar la variabilidad de la misma o cuando es posible una mayor certeza en el rendimiento sin disminuirlo. Markowitz plantea el problema como un programa de optimización de la rentabilidad esperada de la cartera para cada uno de los valores posibles del riesgo de ésta<sup>28</sup>, tal y como se muestra a continuación:

---

<sup>28</sup> En la actualidad existen programas informáticos estándar que nos proporcionan la solución a este problema de programación matemática.

$$\max E(\tilde{R}_p) = \sum_{i=1}^N w_i E(\tilde{R}_i)$$

Sujeto a:

$$\sigma_p^2 = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} = V^*$$

$$\sum_{i=1}^N w_i = 1$$

$$w_i \geq 0$$

Donde  $V^{*29}$  es un valor fijado de antemano para la varianza del rendimiento de la cartera. La aplicación de este programa nos permite determinar en qué activos invertir y en qué proporciones,  $w_i$ , para que la rentabilidad esperada sea máxima dado un nivel de riesgo,  $V^*$ . Si repetimos este proceso para los distintos valores de riesgo se genera la frontera eficiente.

**C) Especificación de las curvas de indiferencia del inversor.** Cualquier inversor racional en situación de riesgo intentará siempre maximizar su utilidad esperada de su riqueza. Aunque cada inversor tiene su propia función de utilidad específica y particular, dado que Markowitz supone que los inversores son adversos al riesgo, sus funciones de utilidad serán estrictamente crecientes y cóncavas.

---

<sup>29</sup> El asterisco denota que el parámetro correspondiente,  $V$  para la varianza y  $E$  para el caso de la rentabilidad esperada de la cartera, puede adoptar diferentes valores.

El criterio media varianza no nos permite establecer comparaciones entre dos carteras eficientes, pues la cartera más arriesgada a la vez ofrece mayor rendimiento esperado que la otra. Así pues, sólo será posible conocer la elección de la cartera óptima, entre todo el conjunto de carteras eficientes, si sabemos cuál es el grado concreto de aversión al riesgo que presenta el inversor, es decir, cuáles son sus preferencias.

Por tanto, la función de utilidad de un inversor, medida en términos del rendimiento esperado y del riesgo de inversión, se traduce en una determinada relación marginal de sustitución o de intercambio entre esos dos elementos que conforman un mapa de curvas de indiferencia.

Una curva de indiferencia o isoutilidad está representada por el conjunto de combinaciones rendimiento esperado-riesgo que proporcionan al inversor la misma utilidad esperada. De esta forma, la pendiente de una curva de indiferencia en uno de sus puntos indica el aumento en el rendimiento esperado que el individuo exige a la inversión por unidad de riesgo adicional, o también el incremento de riesgo que está dispuesto a asumir si puede obtener un incremento unitario en la rentabilidad esperada, de modo que la utilidad esperada se mantenga constante.

A lo largo de cada curva de indiferencia, el nivel de utilidad esperado obtenido por el individuo es el mismo. Respecto a las curvas de indiferencia se pueden hacer las siguientes consideraciones:

Las curvas de indiferencia de cualquier inversor racional y averso al riesgo son crecientes y convexas, la forma de especificar dichas curvas dependerá de la actitud particular, que puede ser más o menos arriesgada.

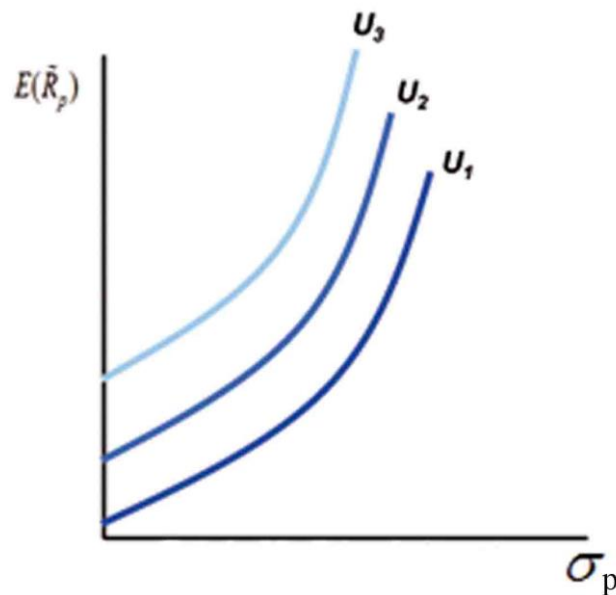


Figura 2.2. Curva de indiferencia de un inversor muy averso al riesgo.

Fuente: elaboración propia.

Así, en las figuras 2.2 y 2.3 se representan las curvas de indiferencia de dos inversores racionales y adversos al riesgo. Puede observarse que el grado de aversión al riesgo del inversor cuyas curvas de indiferencia se representan en la figura 2.2, es mayor que el inversor correspondiente a la figura 2.3, ya que aquél exige un incremento más elevado de rentabilidad esperada para compensar un aumento unitario de riesgo.

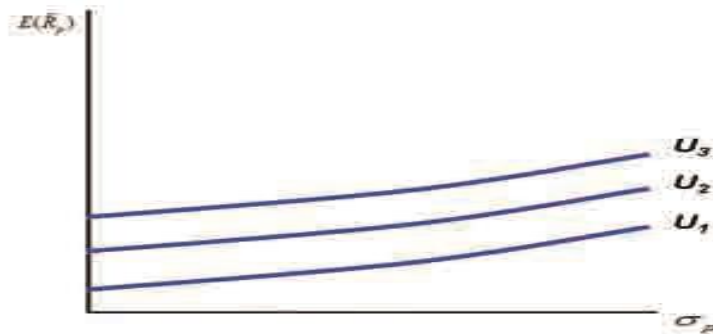


Figura 2.3. Curva de indiferencia de un inversor poco adverso al riesgo.

Fuente: elaboración propia.

Las curvas de indiferencia no se pueden cortar. Si observamos la figura 2.4, en donde las curvas  $U_1$  y  $U_2$  se cortan, vemos que los puntos A y B de la curva de indiferencia  $U_1$  proporcionan al inversor la misma utilidad esperada. Asimismo, los puntos B y C de la curva de indiferencia  $U_2$  también proporcionan al inversor la misma utilidad esperada. De este modo, tendríamos que A es indiferente a B y B es indiferente a C y, suponiendo que las preferencias son transitivas, A es indiferente a C, lo cual no es válido ya que el punto A es preferido al C pues tiene un rendimiento esperado superior con idéntica varianza.

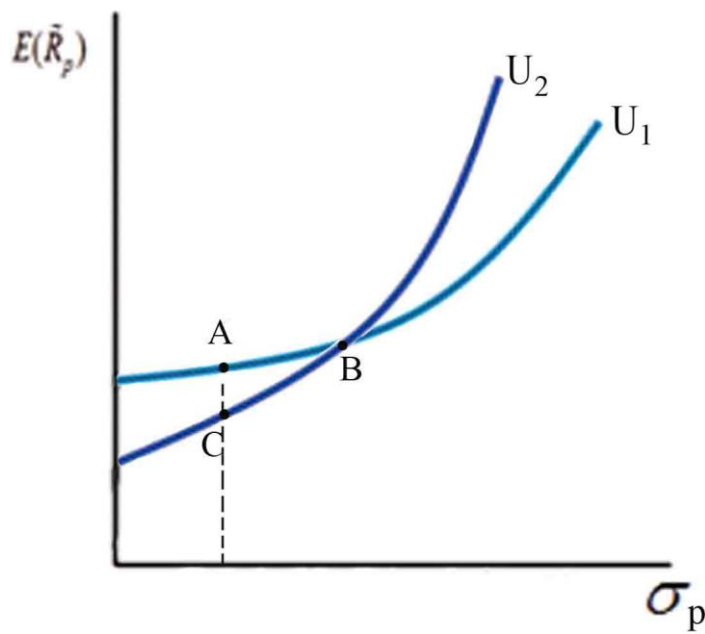


Figura 2.4.

Fuente: elaboración propia.

- Una curva de indiferencia, en su corte con el eje de ordenadas, nos proporciona el equivalente cierto de cualquier inversión arriesgada contenida en ella, esto es, aquel rendimiento cierto para el cual el individuo se muestra indiferente entre dicho rendimiento y aquel otro rendimiento de una inversión arriesgada (con varianza positiva) contenida en la curva de indiferencia en cuestión.
- Como suponemos inversores racionales, las curvas de indiferencia no se prolongan por debajo del eje de abscisas, ya que representaría inversiones con rendimientos esperados negativos y varianzas positivas que un inversor racional no contempla.



- Las curvas de indiferencia situadas más hacia arriba y más hacia la izquierda representan niveles de utilidad esperada superiores, ya que el rendimiento esperado para un mismo riesgo o varianza es mayor.

**D) Determinación de la cartera óptima del inversor.** Una vez determinados la frontera eficiente y el mapa de curvas de indiferencia de un inversor, estamos en condiciones de obtener su cartera óptima, definida como aquella cartera, entre todas las que configuran la frontera eficiente, correspondiente a la combinación media-desviación típica (o media-varianza) que maximiza su utilidad esperada, esto es, aquella cartera que mejor se ajusta a sus preferencias personales sobre el riesgo.

Unos inversores preferirán una ganancia mayor aunque para ello tengan que soportar un riesgo mayor y otros se conformarán con un riesgo menor a cambio de un riesgo también inferior.

Superponiendo la frontera eficiente y el mapa de curvas de indiferencia de un inversor particular se puede determinar cuál es su cartera óptima, es decir, aquella cartera situada sobre una curva de indiferencia lo más arriba y lo más hacia la izquierda posible que le proporcionará al decisor la máxima utilidad esperada.

### 2.1.2. CAPM

El modelo de Markowitz es un modelo normativo, sirve de guía de acción al inversor que intenta maximizar su utilidad esperada, le indica en qué activos financieros invertir y con qué peso y ponderación.

De igual importancia que lo anterior es conocer que cabe esperar a nivel de mercado, qué precios de equilibrio de los activos financieros se van a dar en él, cuando todos los inversores toman sus decisiones de inversión financiera de acuerdo con el modelo de los parámetros de selección de carteras. Esta generalización del modelo de Markowitz ha dado lugar al desarrollo de la teoría del mercado de capitales a la que habitualmente, por comodidad, se hace mención con las iniciales inglesas CAPM (Capital Asset Pricing Model).

A diferencia del modelo de Markowitz, el modelo CAPM es un modelo positivo: intenta averiguar las implicaciones positivas a nivel de mercado que se derivan de la teoría de carteras en el caso de que todos los inversores sean diversificadores eficientes en el sentido Markowitz y el mercado cumpla unas determinadas hipótesis.

Según señala Reig Pérez (2006, pp 277-306,) en orden cronológico, la versión más temprana del CAPM se debe a Sharpe (1964). Sin embargo, en la literatura sobre finanzas lo normal es hacer referencia a otros tres autores como responsables del inicio y posterior desarrollo de la teoría del mercado de capitales: Sharpe (1964), Lintner (1965) y Mossin (1966).

Como toda teoría, el modelo CAPM se apoya en una serie de supuestos restrictivos sobre la realidad que intenta explicar. A los supuestos relativos al comportamiento de los inversores y a las características de los activos y mercado desarrollados por Markowitz, se añaden ahora dos supuestos más sobre los mercados financieros:

- A. Los inversores tienen expectativas homogéneas. Es decir, todos llegan a las mismas estimaciones de  $E(\tilde{R}_i)$ ,  $\sigma(\tilde{R}_i)$  y  $\text{cov}(\tilde{R}_i, \tilde{R}_j)$ , ya que al ser coincidente su punto de vista sobre el curso de la evolución futura de los precios de los títulos, parten de las mismas funciones de distribución de los rendimientos de los mismos.
- B. Los mercados de capitales están en equilibrio al principio del período único de planificación o, lo que es lo mismo, la oferta de los títulos iguala la demanda de los mismos<sup>30</sup>.

#### 2.1.2.1. CML

Si los individuos son diversificadores eficientes en el sentido Markowitz y tiene a su disposición el activo libre de riesgo, la frontera eficiente se vuelve lineal (figura 2.5), invirtiendo el individuo, de acuerdo con el teorema de la separación de Tobin, sólo en dos activos: el activo libre de riesgo F y la cartera T situada en

---

<sup>30</sup> También debe cumplirse que la suma tomada a préstamo por la totalidad de los inversores coincida con la cantidad total por ellos prestada.

el punto de tangencia de la frontera eficiente cóncava (encontrada cuando sólo se invierte en  $N$  activos arriesgados) con la recta que parte del punto  $R_F$ .

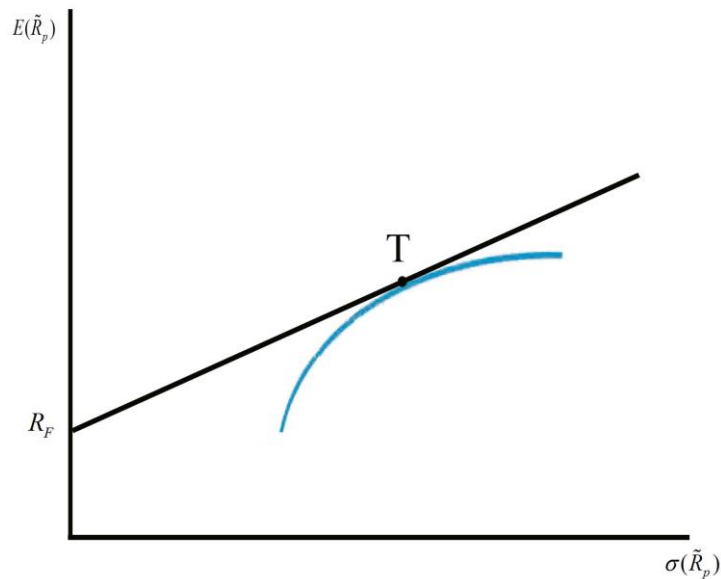


Figura 2.5. Frontera eficiente con  $n$  activos arriesgados más el activo libre de riesgo.

Fuente: elaboración propia.

Puesto que el mercado de capitales está en equilibrio, la única manera de que todos los inversores puedan ver satisfecha la demanda de títulos arriesgados que refleja la composición de la cartera  $T$  es que dicha demanda coincida con la oferta global que de dichos títulos hay en el mercado, lo cuál sólo será posible si la cartera tangente  $T$  es la cartera de mercado  $M$  compuesta por los títulos arriesgados existentes en el mercado en las proporciones que representa el valor de mercado de dichos títulos respecto al valor total del conjunto de todos los títulos en el mercado (figura 2.6).

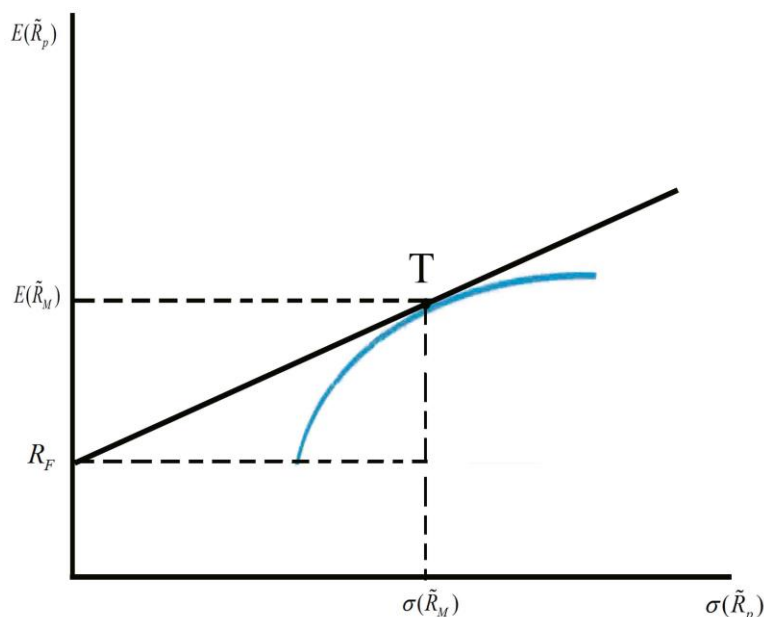


Figura 2.6. Línea del mercado de capitales (CML).

Fuente: elaboración propia.

La cartera M, que desean mantener con mayor o menor peso todos los inversores, es independiente de la actitud que cada inversor tenga frente al riesgo, siendo su composición la misma que la del mercado. Si denominamos  $P_i$  y  $Q_i$  al precio y al número de títulos de cada uno de los títulos arriesgados negociados en el mercado, la proporción con que interviene cada valor dentro de la cartera de mercado será el porcentaje que representa la capitalización de cada valor sobre la capitalización total del mercado:

$$w_i^M = \frac{P_i Q_i}{\sum_{i=1}^N P_i Q_i} \quad \text{siendo:} \quad \sum_{i=1}^N w_i^M = 1$$

A la recta que representa la frontera eficiente de todos los inversores (figura 2.6), y que es consecuencia directa de la introducción de los supuestos de

las expectativas homogéneas y el equilibrio de mercado, la denomina línea de mercado de capitales (Capital Market Line), dependiendo su ordenada en el origen ( $R_F$ ) y la pendiente  $((E(\tilde{R}_M) - R_F) / \sigma_M)$  únicamente de dos activos financieros: uno sin riesgo (cuya rentabilidad es  $R_F$ ) y otro con riesgo (con tasa de rentabilidad  $\tilde{R}_M$ ). Analíticamente su expresión es:

$$E(\tilde{R}_p) = R_F + \left( \frac{E(\tilde{R}_M) - R_F}{\sigma_M} \right) \sigma_p$$

Si llamamos  $\lambda$  a la constante siguiente:

$$\lambda = \left( \frac{E(\tilde{R}_M) - R_F}{\sigma_M} \right)$$

La ecuación de la CML también puede escribirse del siguiente modo:

$$E(\tilde{R}_p) = R_F + \lambda \sigma_p$$

Siendo  $E(\tilde{R}_p)$  y  $\sigma_p$  medidas adecuadas para medir la rentabilidad y el riesgo de una cartera de valores en los ejes  $(E, \sigma)$ , por lo que la CML nos indica la relación entre la rentabilidad y el riesgo de las carteras eficientes.

Sólo las carteras eficientes se sitúan sobre la CML pudiéndose descomponer la rentabilidad que proporcionan dichas carteras en dos partes:

$R_F$ , la ordenada en el origen de la CML, que es la tasa pura de interés o precio del tiempo y representa el premio por esperar y consumir más tarde, en lugar de hacerlo en el momento presente.

$\lambda$ , la pendiente de la CML, que es una constante que multiplica al riesgo de la cartera y, por ello, recibe el nombre de precio de mercado del riesgo, o simplemente precio del riesgo. Con lo que para aumentar la rentabilidad esperada de una cartera eficiente hay que aceptar un mayor riesgo<sup>31</sup>.

#### 2.1.2.2. SML

Ya que la CML proporciona la relación de equilibrio entre la rentabilidad y el riesgo de carteras eficientes, sería conveniente contar también con una relación similar que proporcione la relación de equilibrio entre la rentabilidad y el riesgo de cualquier cartera (eficiente o no) y de cualquier título. Esa relación es la línea de mercados de títulos (SML o Security Market Line, en inglés) e indica la rentabilidad que por término medio el mercado va a ofrecer a cualquier inversión financiera, tanto si el inversor ha diversificado bien el riesgo como si ha renunciado por completo a cualquier tipo de diversificación y ha colocado todo su dinero en un título.

---

<sup>31</sup> Es importante señalar que la rentabilidad de una cartera eficiente es una media esperada con determinada desviación típica, no una rentabilidad cierta sin más. Quiere esto decir que, en un año en particular, la rentabilidad efectiva de una cartera arriesgada puede ser menor que la rentabilidad efectiva con menor riesgo. Sólo en términos medios, la rentabilidad de una cartera eficiente más arriesgada es también mayor.

La expresión analítica de la SML es:

$$E(R_i) = R_F + (E(\tilde{R}_M) - R_F)\beta_i$$

siendo

$$\beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$$

En donde el subíndice i puede indicar bien un título individual o bien una cartera, eficiente o no.

Habitualmente, una expresión alternativa que se suele utilizar con frecuencia para hacer referencia a la ecuación de la recta SML es la siguiente:

$$E(\tilde{R}_i) = R_F + \lambda' \beta_i$$

En donde  $\lambda'$ , esto es:  $E(\tilde{R}_M) - R_F$ , recibe el nombre de “Risk Premium” del mercado o prima por asumir más riesgo. Gráficamente:



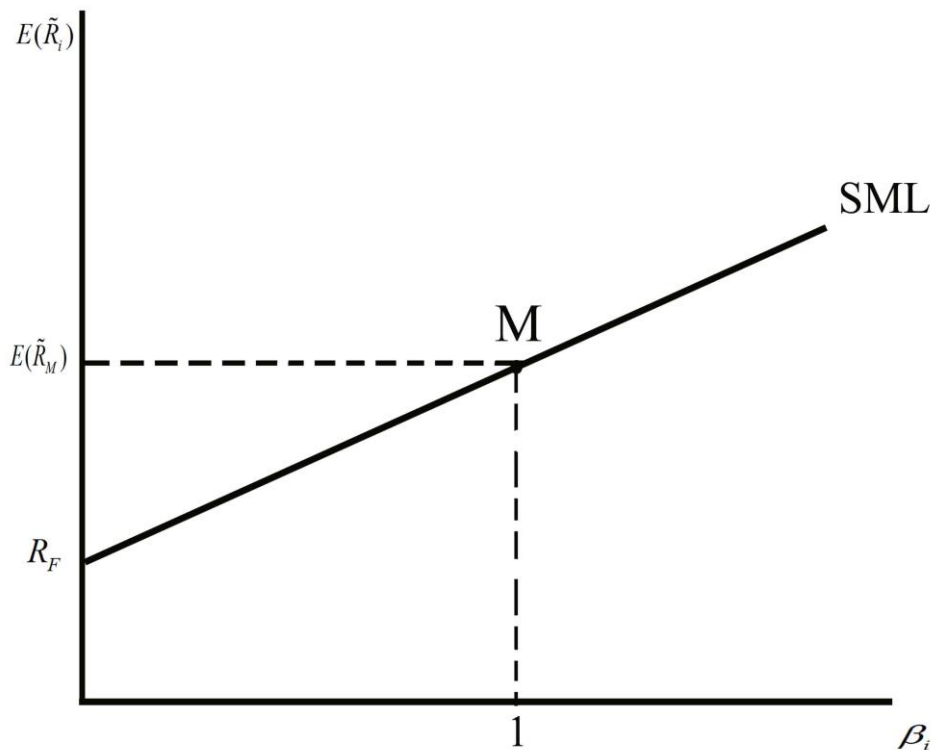


Figura 2.7. Línea del mercado de títulos (SML).

Fuente: elaboración propia.

Nótese que la beta de una cartera de mercado es la unidad<sup>32</sup>, ya que si se regresa  $\tilde{R}_M$  frente a  $\tilde{R}_M$ , la relación lineal es la recta de 45 grados.

Los títulos individuales o carteras se pueden clasificar, con base en su coeficiente beta en 3 tipos:

- Agresivos:  $\beta_i > 1$ . Son títulos recomendados para una bolsa alcista. Las rentabilidades de esos títulos serán mayores a la rentabilidad del mercado.

<sup>32</sup> La beta del título libre de riesgo es nula. Esto, que se puede visualizar en la gráfica de la SML, se debe a que, al tratarse de una constante, la desviación típica de  $R_F$ , así como la varianza de  $R_F$  con  $\tilde{R}_M$  son nulas.

- Neutros:  $\beta_i = 1$ . Estos títulos suben y bajan al mismo ritmo que el mercado, ya que al tener la misma beta que él, su rentabilidad esperada es la misma.
- Defensivos:  $\beta_i < 1$ . En épocas alcistas, estos títulos suben con menos intensidad que lo hace el mercado; aunque también es cierto que en épocas bajistas provocan menos pérdidas que la cartera de mercado, por lo que se les califica de defensivos.

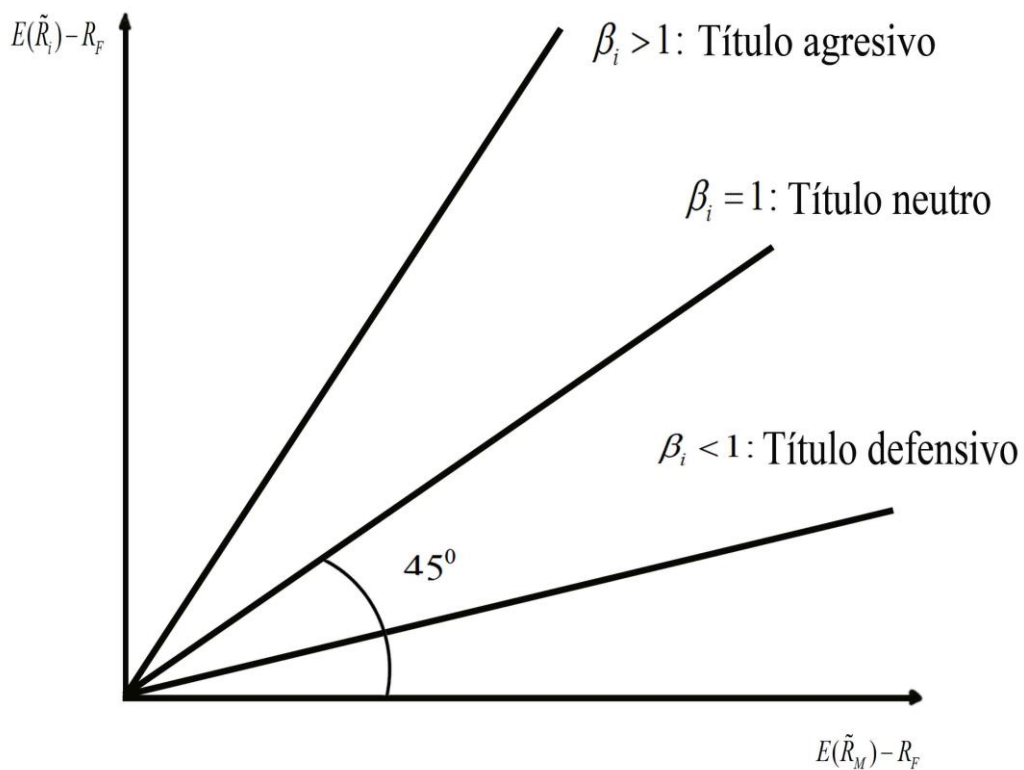


Figura 2.8.

Fuente: elaboración propia.

La línea del mercado de capitales y la línea característica de títulos indican relaciones de equilibrio. Si las relaciones no se cumplen, aparecerán presiones

tendientes al cambio. Sharpe<sup>33</sup> llega a demostrar que su modelo cumple las condiciones exigidas por los economistas para que su modelo sea considerado de equilibrio es decir; que es estable, que es único y estable.

Las conclusiones que Sharpe<sup>34</sup> obtiene de su modelo en condiciones de concordancia son las siguientes:

1. La línea de mercado de capitales, proporciona una representación gráfica de las nociones de precio del riesgo y precio del consumo inmediato. También mide la eficacia de una cartera.
2. La línea de mercado de títulos aporta una representación gráfica de la relación entre rentabilidad esperada y riesgo de los títulos. El concepto de riesgo considera la existencia de diversificación y el deseo de realizarla. El concepto de riesgo sistemático, nos aclara la diferencia entre la incertidumbre y el riesgo. Éste último es diversificable, la primera no lo es en gran medida.
3. Las líneas características relacionan, los tipos de rentabilidad de los títulos o de las carteras con el de la cartera de mercado. Se pueden relacionar estas líneas con el concepto de eficacia. Una cartera es eficaz si (y sólo si) su línea característica integra toda la incertidumbre relacionada con su tipo de rentabilidad.

---

<sup>33</sup> Ver Sharpe, W.F., Teoría de cartera y del mercado de capitales, Deusto, 1.975, anexos sección D y pp 127-130.

<sup>34</sup> Ver Sharpe, W.F., Teoría de cartera y del mercado de capitales, Deusto, 1.975, pp 130-132.

4. Las carteras eficaces se construirán por combinaciones de la cartera de mercado y de créditos y préstamos. Esto entra en contradicción con la conducta observada en la realidad. Para poder admitirlo tenemos que retirar el concepto de concordancia absoluta, dando lugar a otro modelo denominado de discordancia. Otra posibilidad es atribuir este comportamiento a los costes de transacción y de gestión que harían que los inversores escogieran carteras eficaces.
5. Se trata de una aplicación práctica limitada. Estrictamente hablando es incorrecto creer que la línea del mercado de capitales es un conjunto de opciones presentado al inversor. Igualmente, es incorrecto creer que un título está sobreponderado si se representa debajo de la línea del mercado de títulos. Esto sólo se cumpliría una vez alcanzado el equilibrio, mientras que las anteriores conclusiones se utilizan en la práctica de modo contrario a su sentido cuando la persona todavía tiene que tomar sus decisiones.

El modelo de discordancia se diferencia del anterior solamente en un aspecto: se admite que no existe un consenso sobre el futuro<sup>35</sup>. Si se admite la discordancia<sup>36</sup> el tipo de interés nominal es el único denominador común, todo lo demás es subjetivo. No existe una sola combinación óptima de títulos con riesgo; la combinación preferida dependerá de las preferencias del inversor. En síntesis, el teorema de separación no es válido. La teoría ya no es válida, la eliminación de la teoría del mercado de capitales va seguida inmediatamente por la línea del

---

<sup>35</sup> Ver Sharpe, W.F., Teoría de cartera y del mercado de capitales, Deusto, 1975, pp 134.

<sup>36</sup> Ver Sharpe, W.F., Teoría de cartera y del mercado de capitales, Deusto, 1975, pp 134-143.

mercado de títulos. Este modelo explica mejor la realidad pero, al explicar todo no explica nada. Sus deducciones son lo suficientemente generales como para ser coherentes con casi cualquier experiencia observada. No es un modelo de uso en la práctica, a no ser que el inversor crea que puede efectuar predicciones más afinadas sobre el futuro que los demás inversores. En este sentido justifica la posibilidad de realizar el análisis bursátil. El modelo anterior es adecuado para inversores que carecen de buena información<sup>37</sup>. No tener información buena, es una situación equivalente a que todos tengan la misma información. Asumir que todos los inversores tiene la misma información es un hecho asimilable a decir que nadie tiene mejor información que los demás, y en ese sentido nadie tiene una información “buena” que le permita obtener rendimientos por encima del resto del mercado. Las únicas diferencias en los rendimientos se deben al distinto riesgo de los activos, esto es lo que dice el modelo de concordancia. También es lo que dice la “hipótesis de eficiencia de los mercados” o del “mercado eficiente” y es la razón para no considerarla una teoría diferente de la que implica el modelo de concordancia. El modelo de discordancia niega esa posibilidad, y considera el análisis bursátil válido.

### **2.1.3. Modelo de Mercado**

Sharpe continuó los estudios sobre la teoría de carteras, iniciados por Markowitz, ampliándolos y llegando a encontrar una fórmula para su aplicación práctica, además de desarrollar una teoría de equilibrio del mercado de capitales que le permitió valorar de forma lógica los activos.

---

<sup>37</sup> Ver Sharpe, W.F., Teoría de cartera y del mercado de capitales, Deusto, 1975, pp 143.

William Sharpe, uno de los doctorandos a los que Markowitz impartía clase, amplió los estudios iniciados por su maestro en el área de la teoría de carteras. El problema que pretende solucionar, es reducir el elevado número de estimaciones que hay que realizar para utilizar el modelo de Markowitz, lo que lo invalida para su utilización en la práctica<sup>38</sup>.

El denominado modelo de mercado de Sharpe, convenientemente matizado por Treynor<sup>39</sup> nos dice que la rentabilidad de cualquier título o cartera depende de tres componentes (  $\alpha, \beta, \tilde{\varepsilon}$  ). Por otra parte, permite caracterizar el riesgo de una cartera a efectos de calibrar la remuneración merecida por el inversor distinguiendo entre riesgo sistemático y riesgo no sistemático y admite el análisis de oportunidades de arbitraje.

#### 2.1.3.1. Rentabilidad de un título

El modelo de mercado afirma que la rentabilidad de cualquiera de los diferentes títulos con riesgo existentes en una economía (en un número igual a N) a lo largo de un determinado período de tiempo t (por ejemplo, la del título

---

<sup>38</sup> El número total de estimaciones a realizar en el modelo de Markowitz se obtiene de la expresión  $(N^2 + 3N)/2$ , siendo N el número de títulos de la muestra considerada. La incorporación de un nuevo título a la cartera supone realizar N+2 estimaciones adicionales. Por ejemplo, una cartera con 500 títulos serían necesarios 125.750 estimaciones.

<sup>39</sup> Sharpe, W.F., A Simplified Model for Portfolio Analysis, "Management Science", Vol IX, nº2 enero, 1.963, pp. 277-293.; J.L. Treynor: How to rate management of Investment Funds, "Harvard Business Review", enero-febrero 1965, pp. 63-75.

$i: \tilde{R}_{it}, \forall i=1, \dots, N$ ) es una variable aleatoria que depende linealmente de tres componentes, a saber:

1. El primero,  $\alpha_{it}$ , refleja la parte de  $\tilde{R}_{it}$  debida a todas aquellas causas que, durante el período considerado, afectan a la rentabilidad del título y que no pueden ser explicadas por la evolución general de la economía pero quizá sí por ciertas características intrínsecas no fortuitas (y por tanto predecibles) de la empresa emisora del título. A este término, cuyo valor (mayor o menor que 0) se supone constante a lo largo del tiempo lo denominaremos coeficiente alfa o simplemente alfa del título.
2. El segundo hace referencia a la parte de  $\tilde{R}_{it}$  debida a aquellas causas que afectan a la rentabilidad del título y que pueden ser explicadas por la evolución general de la economía a lo largo del período analizado. Su influencia sobre la rentabilidad de un título se designa mediante el producto  $\beta_{it} \tilde{R}_{Mt}$ , de forma que, dado un valor de  $\tilde{R}_{Mt}$ , cuanto mayor sea  $\beta_{it}$  (en módulo) mayor será esa influencia (y viceversa). A este término se le denomina componente beta o la beta del título i. El valor de dicho coeficiente deber ser, normalmente, positivo ( $\beta_{it} > 0$ ), si bien es posible teóricamente la existencia de títulos contracíclicos, esto es, cuya rentabilidad evolucione en sentido inverso a la de la cartera de mercado ( $\beta_{it} < 0$ )
3. El tercero (variable aleatoria  $\tilde{\varepsilon}_{it}$ ) recoge una serie indeterminada de efectos particulares (sobre  $\tilde{R}_{it}$ ) provocado por causas intrínsecas fortuitas

(impredecibles) favorables ( $\varepsilon_{it} > 0$ ) y desfavorables ( $\varepsilon_{it} < 0$ ) que, en su caso, afectan a la empresa emisora del título en cuestión en algún o algunos momentos a lo largo del período considerado, independientemente de la marcha general de la economía durante dicho período. Lo denominaremos componente épsilon o perturbación aleatoria del título.

Sharpe, por tanto, sostiene que las rentabilidades de los valores, están relacionadas por medio de su dependencia a un factor subyacente que es común a todos. En mayor o menor grado, la rentabilidad de cada título depende de uno o más índices, tales como: el producto nacional bruto, la renta per cápita o un índice bursátil. En su artículo<sup>40</sup> se centra en un solo índice, aunque en el libro<sup>41</sup> publicado con posterioridad dedica todo un capítulo al estudio de éste y otros casos donde se utilizan múltiples índices.

$$\tilde{R}_{it} = \alpha_{it} + \beta_{it} \tilde{R}_{Mt} + \tilde{\varepsilon}_{it}$$

$$i=1, \dots, N$$

$$t=1, \dots, T$$

- N: número de títulos con riesgo en la economía.
- T: número de períodos considerados.

Siendo durante el período t:

---

<sup>40</sup> Sharpe, W.F., A Simplified Model for Portfolio Analysis, "Management Science", Vol IX, nº2 enero, 1.963, pp. 277-293.

<sup>41</sup> Sharpe, W.F., Teoría de carteras y del mercado de capitales, Deusto, 1.975.



- $\tilde{R}_{it}$  : rentabilidad (aleatoria) del título i.
- $\alpha_{it}$  : coeficiente alfa del título i.
- $\beta_{it}$  : coeficiente beta del título i.
- $\tilde{R}_{Mt}$  : rentabilidad (aleatoria) de la cartera de mercado.
- $\tilde{\varepsilon}_{it}$  : perturbación aleatoria del título i.

El rendimiento de un título se explica en parte por el rendimiento del mercado y otros factores propios del título. La incidencia macroeconómica en el modelo es, por tanto, evidente: cuanto más favorable sea la evolución prevista de la economía (y por tanto de las magnitudes macroeconómicas que la describen) mayor será la rentabilidad esperada del mercado, siendo  $\beta_i > 0$ , y mayor deberá ser a su vez la rentabilidad esperada del título.

Es preciso hacer notar que el modelo de mercado no proporciona una relación funcional entre la rentabilidad y el riesgo de un título (que es realmente lo que, en última instancia, precisa el inversor), sino entre la rentabilidad de dicho título y la rentabilidad de la cartera de mercado.

#### 2.1.3.2. Rentabilidad de una cartera

En virtud del principio de diversificación, un inversor racional no dedicará el 100 por 100 de su presupuesto de inversión a un único título, sino que distribuirá dicho presupuesto (agotándolo) entre un número determinado de títulos formando una cartera. El modelo de mercado aplicado a una cartera (que

denominamos P) compuesta por un número H de títulos con riesgo (número no necesariamente igual al total de los títulos existentes:  $H \leq N$ ) tendrá la siguiente forma, en la que el coeficiente alfa, el coeficiente beta y la perturbación aleatoria de la cartera serán, en cada caso, una media ponderada de los coeficientes alfa, beta y de la perturbación aleatoria, respectivamente, de cada uno de los títulos que la integran, coincidiendo la correspondiente perturbación con la proporción del presupuesto de inversión ( $w_i^p$ ) dedicada a cada título.

$$\begin{aligned}\tilde{R}_{pt} &= \sum_{i=1}^{H \leq N} w_i^p \tilde{R}_{it} = \sum_{i=1}^{H \leq N} w_i^p (\alpha_{it} + \beta_{it} \tilde{R}_{Mt} + \tilde{\varepsilon}_{it}) = \\ &= \alpha_{pt} + \beta_{pt} \tilde{R}_{Mt} + \tilde{\varepsilon}_{pt}\end{aligned}$$

para  $t=1, \dots, T$

Siendo durante el período t:

- $w_i^p$ : proporción del presupuesto de inversión dedicado al título i en la cartera P.
- $\tilde{R}_{pt}$ : rentabilidad (aleatoria) de la cartera P.
- $\alpha_{pt} = \sum_{i=1}^{H \leq N} w_i^p \alpha_{it}$ : coeficiente alfa de la cartera P.
- $\beta_{pt} = \sum_{i=1}^{H \leq N} w_i^p \beta_{it}$ : coeficiente beta de la cartera P.
- $\tilde{\varepsilon}_{pt} = \sum_{i=1}^{H \leq N} w_i^p \tilde{\varepsilon}_{it}$ : perturbación aleatoria de la cartera P.

La rentabilidad de cualquier cartera está, de la misma forma que la de cualquier título, cualitativamente afectada por los tres componentes (alfa, beta y épsilon). Una cartera P se diferenciará de otra P' igualmente integrada por el mismo número (H) y tipo de títulos con riesgo que aquella en el hecho de que a alguno de tales títulos se le habrá dedicado diferente proporción del presupuesto de inversión ( $w_i^P \neq w_i^{P'}$ ); en consecuencia, la rentabilidad de dichas carteras diferirá.

El modelo de mercado no proporciona una relación funcional entre la rentabilidad y el riesgo de la cartera sino entre la rentabilidad de dicha cartera y de la cartera de mercado.

La aplicación del modelo de estimación mínimo-cuadrático exige, como es sabido, un conjunto de hipótesis previas con relación a la perturbación aleatoria que, en el ámbito del modelo de mercado y para el caso del título i, son las siguientes:

1. En cualquier período t los distintos factores que, en forma de perturbaciones aleatorias, afecten, en su caso, a la rentabilidad de un título tiene un efecto esperado conjunto nulo sobre la rentabilidad esperada del mismo (los efectos positivos y los negativos se compensan mutuamente en su globalidad):

$$E[\tilde{\varepsilon}_{it}] = 0 \quad , \quad \forall t$$

2. En cualquier período  $t$  la varianza de cualquier perturbación aleatoria, que, en su caso, afecte a la rentabilidad de un título determinado es la misma (esto es, tiene un valor constante igual a  $\sigma_{\tilde{\varepsilon}_t}^2$ ):

$$\sigma_{\tilde{\varepsilon}_t}^2 = \sigma_{\tilde{\varepsilon}_i}^2, \quad \forall t$$

3. Cualquier perturbación aleatoria que, en su caso, afecte a la rentabilidad de un título determinado no dependen en modo ni período alguno de la evolución de la economía (representada por la evolución de la cartera  $M$ ); por lo tanto la covarianza entre ambas variables es nula:

$$\sigma_{\tilde{\varepsilon}_t \tilde{R}_{Mt}}^2 = 0, \quad \forall t$$

4. Cualquier perturbación aleatoria que, en su caso, afecte a la rentabilidad de un título determinado a lo largo de un período de tiempo no influye ni está influida por cualquier otra que pueda afectar a dicho título en otro período distinto:

$$\sigma_{\tilde{\varepsilon}_t \tilde{\varepsilon}_{t'}}^2 = 0, \quad \forall t \neq t'$$

5. Cualquier perturbación aleatoria que, en su caso, afecte a la rentabilidad de un título determinado a lo largo de un período de tiempo no influye ni está influida por cualquier otra que pueda afectar a dicho título en cualquier período:

$$\sigma_{\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{it'}}^2 = 0 \quad , \quad i \neq j, \forall t$$

6. Cualquier perturbación aleatoria que afecte a la rentabilidad de un título determinado sigue en cualquier período una distribución normal; por tanto, se tratará de una normal de media cero y varianza constante igual a  $\sigma_{\tilde{\varepsilon}_i}^2$  (hipótesis de normalidad):

$$\tilde{\varepsilon}_{it} \rightarrow N[0, \sigma_{\tilde{\varepsilon}_i}^2] \quad , \quad \forall t$$

#### 2.1.3.3. Riesgo de un título

Además de la rentabilidad que espera obtener, el inversor debe manejar otra variable fundamental cual es el riesgo derivable de su inversión. Dado que el modelo de mercado está basado en la estructura propia de un modelo media-varianza, parece lógico utilizar la varianza para medir el riesgo. El riesgo de un título tiene la siguiente expresión:

$$\begin{aligned} \sigma_{\tilde{R}_{it}}^2 &= \sigma_{\alpha_{it} + \beta_i + \tilde{\varepsilon}_{it}}^2 = \beta_i^2 \sigma_{\tilde{R}_{Mt}}^2 + \sigma_{\tilde{\varepsilon}_{it}}^2 + 2\beta_i \sigma_{\tilde{R}_{Mt}\tilde{\varepsilon}_{it}}^2 = \\ &= \beta_i^2 \sigma_{\tilde{R}_{Mt}}^2 + \sigma_{\tilde{\varepsilon}_{it}}^2 \end{aligned}$$

Siendo  $\sigma_{\tilde{R}_{Mt}\tilde{\varepsilon}_{it}}^2 = 0$ .

Como puede observarse, el riesgo del título  $i$  se descompone en dos sumandos:

1. El primero ( $\beta_i^2 \sigma_{\tilde{R}_M}^2$ ) indica la parte del riesgo del título derivado de las oscilaciones de la actividad económica global; tales oscilaciones son, obviamente, intrínsecas al funcionamiento de la economía y se plasman en la variabilidad de la rentabilidad de la cartera de mercado (o en la de la cartera índice que emula) respecto a su valor esperado (representado por la varianza de ésta,  $\sigma_{\tilde{R}_M}^2$ ). El hecho de que la evolución de la economía sea más o menos favorable (o desfavorable) afectará a la variabilidad de la rentabilidad de todos los títulos con riesgo existentes en la misma, si bien no es aventurado pensar que el grado de afección será diferente según el título que en cada caso se analice. Dicho grado de afección queda reflejado en el cuadrado del valor del coeficiente beta ( $\beta_i^2$ ). Por tal motivo, al parámetro  $\beta_i$  (o a su estimación) suele denominársele coeficiente de volatilidad del título  $i$ , aunque, su denominación coloquial es la de “la beta del título  $i$ ”.

A este primer componente se le denomina riesgo sistemático (motivado por el funcionamiento del sistema económico y presente sistemáticamente dado que no es posible eliminar) o de mercado (representado por la evolución del mercado de títulos con riesgo –cuyo índice es precisamente la rentabilidad de la cartera M-) o no diversificable (puesto que no es eliminable mediante el proceso de diversificación).

2. El segundo sumando ( $\sigma_{\tilde{\varepsilon}_{it}}^2$ ) es indicativo del riesgo motivado por causas totalmente ajenas a la incertidumbre en la evolución de la economía. Tales causas deben buscarse en circunstancias particulares que puedan provocar de forma aleatoria cierta variación (favorable o desfavorable y más o menos significativa y/o duradera), en cualquier momento a lo largo del período considerado, en el resultado de la actividad productiva de la empresa y, por tanto, en la rentabilidad que sus títulos puedan finalmente obtenerse en dicho período con respecto a la inicialmente previsible.

Este segundo componente del riesgo es conocido como riesgo no sistemático (puesto que no es causado por el funcionamiento del sistema) o específico (al no ser explicado por la evolución del mercado de títulos con riesgo en su conjunto sino por circunstancias específicas de uno de tales títulos –el que en cada momento sea objeto de análisis-) o diversificable (puesto que es reducible o eliminable al combinar entre sí títulos –bien de forma ingenua, bien de modo eficiente-).

La suma del riesgo sistemático y del riesgo específico es denominada riesgo total ( $\sigma_{\tilde{R}_{it}}^2$ ). Es importante tener en cuenta que, a fin de poder expresar en la misma escala las dos variables fundamentales de cualquier título (rentabilidad esperada y riesgo), el riesgo total suele calcularse no en términos de varianza, sino de desviación típica (esto es, la raíz cuadrada de aquélla:  $\sigma_{\tilde{R}_{it}} = \sqrt{\sigma_{\tilde{R}_{it}}^2}$ ); de esta forma tanto la rentabilidad esperada como el riesgo se miden en unidades no cuadráticas y por tanto mutuamente comparables: el riesgo expresado en porcentaje, indica el tanto por ciento de desviación (tanto al alza como a la baja)

de la rentabilidad realizada respecto a la esperada, expresada en el mismo porcentaje.

#### 2.1.3.4. Riesgo de una cartera

Todo lo dicho del riesgo de un título es aplicable al riesgo de una cartera.

$$\sigma_{\tilde{R}_{pt}}^2 = \beta_p^2 \sigma_{\tilde{R}_{Mt}}^2 + \sigma_{\tilde{\varepsilon}_{pt}}^2$$

$$\sigma_{\tilde{\varepsilon}_{pt}}^2 = \sum_{i=1}^{H \leq N} (w_i^p)^2 \sigma_{\tilde{\varepsilon}_{it}}^2$$

$$\beta_p^2 = \left[ \sum_{i=1}^{H \leq N} w_i^p \beta_i \right]^2$$

Se observa que, al igual que sucede con un único título, el riesgo total de una cartera P ( $\sigma_{\tilde{R}_{pt}}^2$ ) se descompone en riesgo sistemático ( $\beta_p^2 \sigma_{\tilde{R}_{Mt}}^2$ ) y riesgo específico ( $\sigma_{\tilde{\varepsilon}_{pt}}^2$ ), si bien en el caso de la cartera la magnitud de ambos tipos de riesgo depende tanto de los títulos escogidos como de la proporción del presupuesto inicial dedicado a cada uno de éstos ( $w_i^p$ ,  $i=1, \dots, H$ ). Esta circunstancia es fundamental, e indica que, a la hora de diseñar una cartera eficiente (es decir, carente de riesgo específico:  $\sigma_{\tilde{\varepsilon}_{pt}}^2 = 0$ ), tan importante es seleccionar convenientemente los títulos con riesgo que la componen como distribuir adecuadamente entre los mismos el presupuesto de inversión (esto es, determinar el “peso” de cada título en la cartera).



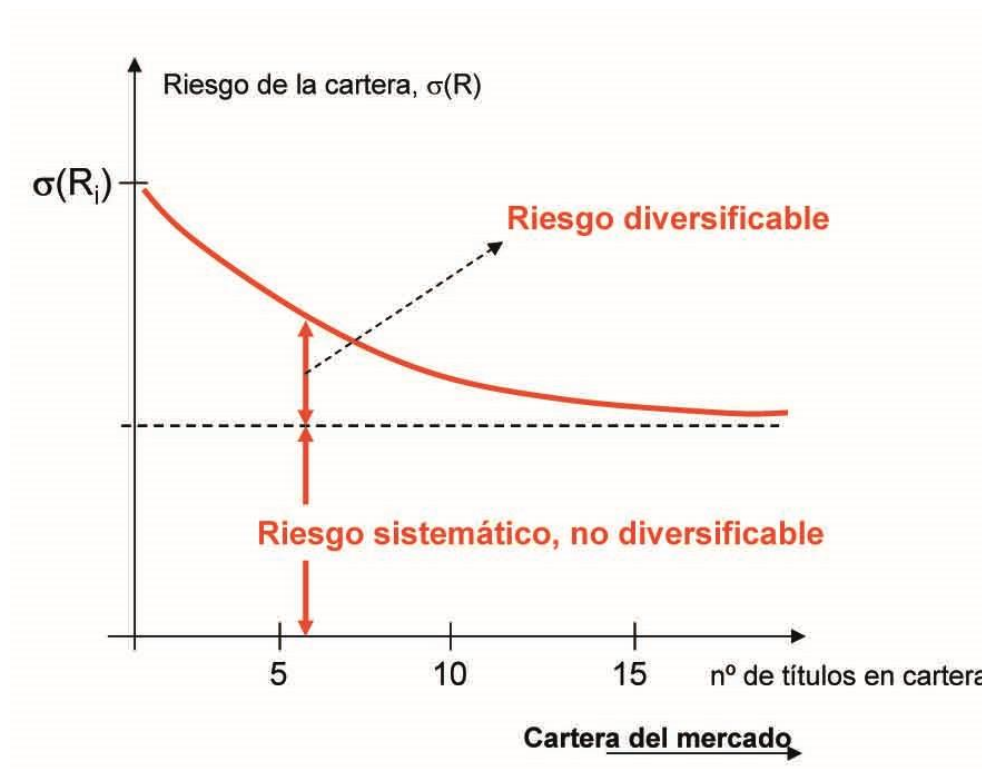


Figura 2.9.

Fuente: elaboración propia

En el eje horizontal se representa el número de títulos que forman parte de la cartera, mientras que en el vertical mide el riesgo de forma relativa frente a esa cartera ideal. La línea curva nos marca el riesgo total asociado con una cartera. Descontando del del valor total, el valor de mercado por la línea recta de puntos tendremos el riesgo específico. La línea de puntos representa la parte del riesgo debida a la incertidumbre de la cartera de mercado.

Sharpe introduce la idea del préstamo y de los activos sin riesgo<sup>42</sup>. Considera necesario para todo el razonamiento la introducción de un título sin riesgo (sinónimo de préstamo), el hecho de poder prestar es independiente de lo anterior. El prestar se admite como una inversión en un título especial que no implica riesgo, en el que su rentabilidad esperada es igual al tipo de interés nominal, porque la desviación típica de su rentabilidad es nula. El pedir crédito puede enfocarse de diversas maneras. En este caso se considera como la emisión de un título carente de riesgo. Si  $x_i$  es el título sin riesgo:

- si  $x_i > 0$ , el inversor presta;
- si  $x_i < 0$ , pide crédito;
- si  $x_i = 0$ , ni presta ni pide crédito.

#### 2.1.4. La Modificación de Treynor

Con objeto de homogeneizar el significado de las variables independientes y dependientes, J.L.Treynor<sup>43</sup> substituyó el índice o cartera de mercado ( $\tilde{R}_{Mt}$ ) por su rendimiento  $\tilde{R}_{Mt} = (I_{t+1} - I_t) / I_t$ .

---

<sup>42</sup> Sharpe llega a la misma conclusión que J.Tobin, aunque por un camino distinto y con posterioridad, que dice que la cartera óptima formada por los activos con riesgo no depende de la actitud frente al riesgo de los inversores sino que es igual para todos ellos. Para lograrlo es necesario introducir en el análisis los títulos sin riesgo. La proposición mencionada es el llamado “teorema de la separación”, demostrado por Tobin en 1.958.

<sup>43</sup> Treynor, J.L., How to Rate Managements of Investments Funds, “Harvard Bussiness Review”, enero-febrero 1965, pp 63-75.

El modelo de Sharpe cobra la siguiente forma:

$$\tilde{R}_{it} = \alpha_{it} + \beta_{it} \tilde{R}_{Mt} + \tilde{\varepsilon}_{it}; \quad t=1,2,\dots,T.$$

En donde

- $\tilde{R}_{it}$  = rendimiento esperado durante el período de tiempo considerado t.
- $\tilde{R}_{Mt}$  = rendimiento del mercado (medido por el rendimiento del índice) durante el mismo período de tiempo. El índice de mercado elegido habrá de referirse a la misma base temporal que el rendimiento del título.
- $\tilde{\varepsilon}_{it}$  = error o perturbación aleatoria.
- $\beta_{it}$  = parámetro que nos indica el peso o intensidad con que las fluctuaciones del rendimiento del mercado afectan al rendimiento del título. Es el llamado coeficiente de volatilidad, riesgo sistemático o de mercado del título i. cuanto mayor sea el coeficiente de volatilidad más aumentará el rendimiento del título i cuando el mercado se halla en alza, pero a su vez será mayor el descenso cuando el mercado esté a la baja. Está en definitiva midiendo el grado de vinculación de la rentabilidad del título i con la rentabilidad del mercado.

- $\alpha_{it}$  = parámetro a estimar que mide la parte de la rentabilidad del título  $i$  que es independiente del mercado. Si la rentabilidad del mercado y el error son cero, el valor de este parámetro coincide con la rentabilidad del título  $i$  cuanto mayor sea  $\alpha_{it}$ , menor vinculación tendrá en principio la rentabilidad de  $i$  con la del mercado.
- $T$  = tamaño de la muestra.

Cuando se verifican las hipótesis convenientes, el método de mercado proporciona las mejores estimaciones de  $\alpha_{it}$  y  $\beta_{it}$ . La recta de regresión que mejor se ajusta a la nube de puntos definida por la ecuación que recoge la simplificación realizada por Treynor se llama línea característica del título  $i$ , que no es más que la recta de regresión que se corresponde con el modelo de mercado de Sharpe, una vez realizado el mencionado cambio de variable al que hemos hecho referencia.

## 2.2. APLICACIÓN A LA ONCOLOGÍA

Podríamos trasladar de la teoría de carteras el estudio del binomio rentabilidad-riesgo, medido mediante los modelos media-varianza, al campo de la oncología molecular. De este modo construiríamos un marco teórico en el que:

La expresión de un gen equivaldría a la rentabilidad de un título.

La cartera de un inversor cualquiera tendría su homólogo en el perfil de expresión de la muestra tumoral<sup>44</sup> correspondiente a un paciente.

### 2.2.1. Expresión de un gen<sup>45</sup>

La expresión génica es el proceso por el cual un gen se activa en una célula<sup>46</sup> para elaborar ARN<sup>47</sup> y proteínas<sup>48</sup>. La expresión de un gen puede medirse al observar el ARN o la proteína elaborada con el ARN, o lo que la proteína hace en una célula.

---

<sup>44</sup> En el capítulo siguiente se describe el proceso mediante el cual se obtiene los perfiles de expresión tanto de las muestras tumorales (CT) como de las muestras de control (CN).

<sup>45</sup> Fragmento de ADN dispuesto en un orden fijo en los cromosomas, que determina la aparición de los caracteres hereditarios en los seres vivos.

<sup>46</sup> La célula es la unidad más pequeña del organismo que dispone de capacidad para actuar de manera autónoma. Uno de los dos tipos de ácido nucleico que elaboran las células.

<sup>47</sup> El ARN contiene información copiada del ADN (el otro tipo de ácido nucleico). Las células elaboran varias formas diferentes de ARN y cada forma cumple una función específica en la célula. Muchas formas de ARN cumplen funciones relacionadas con las proteínas. Para más información véase el Anexo I.

<sup>48</sup> Molécula orgánica rica en nitrógeno, compuesta esencialmente por aminoácidos. Todas las membranas celulares están formadas por proteínas, por lo que constituyen un compuesto fundamental.

No todos los genes se expresan en todas las células todo el tiempo. En realidad, gran parte depende de la capacidad de las células de expresar sus genes en combinaciones, momentos y lugares diferentes. El desarrollo de los organismos multicelulares ofrece un ejemplo asombroso de la denominada “expresión génica diferencial”. Todas las células de un ser humano contienen los mismos genes, pero el conjunto de genes expresados para formar un tipo celular es distinto del expresado para formar los otros. En consecuencia, una célula muscular expresa un conjunto de genes diferente (por lo menos en parte) del expresado por una neurona, una célula epidérmica, etc.

Existen mecanismos que regulan la expresión de un gen, aumentándolo o disminuyéndolo, en función de las necesidades de cada momento.

El proceso, descrito de forma somera es el siguiente: el ADN<sup>49</sup> se encuentra en el núcleo de la célula y está formado por una cadena doble con miles de nucleótidos<sup>50</sup>. Un gen es una secuencia de estos nucleótidos. Al ADN le llegan señales que le indican que tiene que formar ARN. La doble cadena del ADN, momentáneamente se “deshace” y de una de sus hebras se genera otra de ARN (el uracilo se sustituye por la timina). Formada la molécula de ARN, ésta se

---

<sup>49</sup> Abreviatura de ácido desoxirribonucleico (en inglés Deoxyribonucleic Acid o DNA). Es la molécula que contiene y transmite la información genética de los organismos excepto en algunos tipos de virus (retrovirus). Está formada por dos cadenas complementarias de nucleótidos que se enrollan entre sí formando una doble hélice que se mantiene unida por enlaces de hidrógeno entre bases complementarias. Los cuatro nucleótidos que forman el ADN contienen las bases adenina (A), guanina (G), citosina (C) y timina (T). Dado que en el ADN la adenina se empareja sólo con la timina y la citosina sólo con la guanina, cada cadena del ADN puede ser empleada como molde para fabricar su complementaria. Para más información véase el Anexo I.

<sup>50</sup> Molécula constituida por una base nitrogenada, una pentosa y un grupo de ácido fosfórico. Es la unidad básica que compone los ácidos nucleicos. Para más información véase el Anexo I.

traslada al citoplasma<sup>51</sup> donde se localizan otras moléculas como pueden ser aminoácidos, encimas, proteínas, etc. Una vez allí se une al ribosoma<sup>52</sup> para, en función de su secuencia de aminoácidos, sintetizar las proteínas. Mientras las dos hebras de ADN se vuelven a unir. En función del fragmento de ADN seleccionado, intervendrán unos genes u otros y esto hace que las proteínas sintetizadas sean distintas.

Si un gen está sobre-expresado, lo que nos está indicando es que a esa parte concreta del ADN donde se localiza ese gen, le llegan más señales de las habituales para que ponga en marcha, más veces, el mecanismo mediante el cual acaba sintetizando la proteína. Esta proteína a su vez tendrá su función o funciones que, debido a la mayor concentración de aquella verán, seguramente, ampliada su actividad.

Por el contrario, cuando un gen está inhibido hay un déficit de la proteína que codifica. Bien puede ser porque no le lleguen las señales adecuadas para que se inicie el proceso o bien porque el ADN está dañado y aunque le llegan las señales adecuadas para que inicie el proceso, las proteínas que sintetiza no son las correctas debido, por ejemplo, a mutaciones en el ADN lo que implica cambios en la secuencia de bases y por tanto en la proteína final.

Podríamos afirmar que la expresión de cualquiera de los genes que intervienen a lo largo del tiempo (en un número igual a N) en la evolución de un

---

<sup>51</sup> Líquido del interior de la célula, pero exterior al núcleo de la célula. La mayoría de las reacciones químicas de una célula ocurren en el citoplasma.

<sup>52</sup> Orgánulos de las células vivas encargados de la síntesis de las proteínas, compuestos de ácido ribonucleico y proteínas.

tumor (por ejemplo, la del gen  $i: \tilde{R}_{it}, \forall i=1, \dots, N$ ) es una variable aleatoria linealmente dependiente de tres componentes, a saber:

- El primero,  $\alpha_{it}$ , refleja la parte del valor de expresión del gen ( $\tilde{R}_{it}$ ) debida a todas aquellas causas que, durante el período considerado, afectan a la expresión del gen y que no pueden ser explicadas por la evolución general del tumor pero quizá sí por ciertas características intrínsecas no fortuitas (y por tanto predecibles) del propio gen. A este término, cuyo valor (mayor o menor que 0) se supone constante a lo largo del tiempo lo denominaremos coeficiente alfa o simplemente alfa del gen (su homólogo en el análisis bursátil sería el alfa de un título).
- El segundo hace referencia a la parte de la expresión del gen ( $\tilde{R}_{it}$ ) debida a aquellas causas que afectan a la expresión del gen y que pueden ser explicadas por la evolución general del tumor. Su influencia sobre la expresión del gen se designa mediante el producto  $\beta_{it} \tilde{R}_{Mt}$ , de forma que, dado un valor de  $\tilde{R}_{Mt}$ , cuanto mayor sea  $\beta_{it}$  (en módulo) mayor será esa influencia (y viceversa). A este término se le denomina componente beta o la beta del gen i. El valor de dicho coeficiente deberá ser positivo ( $\beta_{it} > 0$ ) cuando el valor de la expresión del gen evolucione en el mismo sentido que la expresión del tumor; y será negativo ( $\beta_{it} < 0$ ) cuando la expresión del gen sea en sentido inverso a la evolución del tumor.



- El tercero (variable aleatoria  $\tilde{\varepsilon}_{it}$ ) recoge una serie indeterminada de efectos particulares (sobre  $\tilde{R}_{it}$ ) provocado por causas intrínsecas fortuitas (impredecibles) favorables ( $\varepsilon_{it} > 0$ ) y desfavorables ( $\varepsilon_{it} < 0$ ) que, en su caso, afectan al gen en cuestión en algún o algunos momentos a lo largo del período considerado, independientemente de la marcha general del tumor. Lo denominaremos componente épsilon o perturbación aleatoria del gen.

Por tanto, asumimos que las expresiones de los genes están relacionadas por medio de su dependencia a un factor subyacente que es común a todos. En mayor o menor grado, la expresión de cada gen depende de una o más variables, tales como: estado de salud del paciente, rutas de señalización, etc.

$$\tilde{R}_{it} = \alpha_{it} + \beta_{it} \tilde{R}_{Mt} + \tilde{\varepsilon}_{it}$$

$$i=1, \dots, N$$

$$t=1, \dots, T$$

- N: número de genes del tumor.
- T: número de períodos considerados.
- Siendo durante el período t:
- $\tilde{R}_{it}$ : expresión (aleatoria) del gen i.
- $\alpha_{it}$ : coeficiente alfa del gen i.
- $\beta_{it}$ : coeficiente beta del gen i.
- $\tilde{R}_{Mt}$ : expresión génica (aleatoria) del tumor.
- $\tilde{\varepsilon}_{it}$ : perturbación aleatoria del gen i.

Es preciso hacer notar que el modelo de mercado, aplicado al campo de la oncología molecular, no proporciona una relación funcional entre la expresión de un gen y su riesgo, sino entre la expresión de dicho gen y la expresión de la nuestra tumoral (cartera de mercado).

### 2.2.2. Riesgo de un gen

Además de la expresión que se espera obtener, debe manejarse otra variable fundamental cual es el riesgo derivable de su variación. Dado que el modelo de mercado está basado en la estructura propia de un modelo media-varianza, parece lógico utilizar la varianza para medir el riesgo. El riesgo de un gen tiene la siguiente expresión:

$$\begin{aligned}\sigma_{\tilde{R}_{it}}^2 &= \sigma_{\alpha_{it} + \beta_i + \tilde{\varepsilon}_{it}}^2 = \beta_i^2 \sigma_{\tilde{R}_{Mt}}^2 + \sigma_{\tilde{\varepsilon}_{it}}^2 + 2\beta_i^2 \sigma_{\tilde{R}_{Mt}\tilde{\varepsilon}_{it}}^2 = \\ &= \beta_i^2 \sigma_{\tilde{R}_{Mt}}^2 + \sigma_{\tilde{\varepsilon}_{it}}^2\end{aligned}$$

Siendo  $\sigma_{\tilde{R}_{Mt}\tilde{\varepsilon}_{it}}^2 = 0$ .

Como puede observarse, el riesgo del gen  $i$  se descompone en dos sumandos:

- El primero ( $\beta_i^2 \sigma_{\tilde{R}_{Mt}}^2$ ) indica la parte del riesgo del gen derivado de las oscilaciones de la actividad tumoral global; tales oscilaciones son, obviamente, intrínsecas al funcionamiento del tumor y se plasman en la variabilidad de la expresión génica del tumor respecto a su valor esperado

(representado por la varianza de ésta,  $\sigma_{\tilde{R}_M}^2$ ). El hecho de que la evolución de la enfermedad sea más o menos favorable (o desfavorable) afectará a la variabilidad de la expresión de la mayoría de los genes (por no decir de todos), si bien no es aventurado pensar que el grado de afección será diferente según el gen que en cada caso se analice. Dicho grado de afección queda reflejado en el cuadrado del valor del coeficiente beta ( $\beta_i^2$ ). Por tal motivo, al parámetro  $\beta_i$  (o a su estimación) le denominaremos coeficiente de volatilidad del gen i, aunque, su denominación coloquial sea la de “la beta del gen i”.

A este primer componente se le denomina riesgo sistemático (motivado por el funcionamiento intrínseco del tumor y que no es posible eliminar) o no diversificable.

- El segundo sumando ( $\sigma_{\tilde{\varepsilon}_{it}}^2$ ) es indicativo del riesgo motivado por causas totalmente ajenas a la incertidumbre en la evolución de la enfermedad. Tales causas deben buscarse en circunstancias particulares que puedan provocar de forma aleatoria cierta variación (favorable o desfavorable y más o menos significativa y/o duradera), en cualquier momento a lo largo del período considerado, como sucede en el resultado de la actividad productiva de la empresa y que incide, por tanto, en la rentabilidad que sus títulos puedan finalmente obtener en dicho período con respecto a la inicialmente previsible.

Este segundo componente del riesgo es conocido como riesgo no sistemático (puesto que no es causado por el funcionamiento del sistema) o específico (al no ser explicado por la evolución del mismo) del o

diversificable (puesto que es reducible o eliminable al combinar entre sí genes).

La suma del riesgo sistemático y del riesgo específico es denominada riesgo total ( $\sigma_{\tilde{R}_{it}}^2$ ). Es importante tener en cuenta que, a fin de poder expresar en la misma escala las dos variables fundamentales de cualquier gen (expresión esperada y riesgo), el riesgo total suele calcularse no en términos de varianza, sino de desviación típica (esto es, la raíz cuadrada de aquélla:  $\sigma_{\tilde{R}_{it}} = \sqrt{\sigma_{\tilde{R}_{it}}^2}$ ); de esta forma tanto la expresión esperada del como el riesgo se miden en unidades no cuadráticas y por tanto mutuamente comparables: el riesgo medido en porcentaje, indica el tanto por ciento de desviación (tanto al alza como a la baja) de la expresión del gen realizada respecto a la esperada, medida en el mismo porcentaje.

### 2.2.3. Perfil de expresión de una cartera

Una cartera, en nuestro modelo, está formada por el conjunto de genes diferencialmente expresados<sup>53</sup> (sobre-expresados o inhibidos) que forman parte del tumor (si se trata de las muestras tumorales), o de las muestras de control (si se trata de las muestras sanas). Cada paciente tiene su propia muestra tumoral, es decir cada paciente tiene su propio perfil de expresión. Todas las carteras (todos los perfiles de expresión de las muestras tumorales y de las muestras de control) están formadas por los mismos genes, y de ahí que la diferencia de la “cartera de un paciente” con la de otro sea la expresión de cada gen. De modo

---

<sup>53</sup> En el estudio se han tomado como diferencialmente expresados un total de 1722 genes.

que el pronóstico de un paciente vendrá indicado por el perfil de expresión de su “cartera”.

Si la “cartera” de un paciente presenta un perfil de expresión malo, esto es, la expresión de los genes que forman dicha cartera no es la adecuada, cabe esperar que el pronóstico para el paciente sea, también, desfavorable. En cambio, sería quizás posible que, cuanto más se acerque el perfil de expresión de una muestra tumoral al perfil de expresión de una muestra sana o muestra de control mejor sea el pronóstico y mayor será la esperanza de vida de dicho paciente.

El modelo de mercado aplicado a una muestra, tumoral o sana, (que denominamos  $P$ ) compuesta por un número  $H$  de genes tendrá la expresión de una recta de regresión donde el coeficiente alfa, el coeficiente beta y la perturbación aleatoria de la cartera serán, en cada caso, una media ponderada de los coeficientes alfa, beta y de la perturbación aleatoria, respectivamente, de cada uno de los genes que la integran.

En condiciones normales, una célula tiene un determinado perfil de expresión, lo que significa que sintetiza un determinado número de proteínas en unas determinadas cantidades (en función de las necesidades de la célula en cada momento). Si la actividad celular cambia, se modificará su perfil de expresión, habrá genes que se expresen más y genes que se expresen menos.

Calcularemos el “peso” o contribución de cada gen en el perfil de expresión de su muestra. Para lo cual sumaremos todos las expresiones de los genes y lo dividiremos entre la expresión de cada gen.

$$\begin{aligned}\tilde{R}_{pt} &= \sum_{i=1}^{H \leq N} w_i^p \tilde{R}_{it} = \sum_{i=1}^{H \leq N} w_i^p (\alpha_{it} + \beta_{it} \tilde{R}_{Mt} + \tilde{\varepsilon}_{it}) = \\ &= \alpha_{pt} + \beta_{pt} \tilde{R}_{Mt} + \tilde{\varepsilon}_{pt} \\ &\text{para } t=1, \dots, T\end{aligned}$$

Siendo durante el período t:

- $w_i^p$ : proporción del presupuesto de inversión dedicado al título i en la cartera de genes P.
- $\tilde{R}_{pt}$ : expresión (aleatoria) de la cartera P.
- $\alpha_{pt} = \sum_{i=1}^{H \leq N} w_i^p \alpha_{it}$ : coeficiente alfa de la cartera P.
- $\beta_{pt} = \sum_{i=1}^{H \leq N} w_i^p \beta_{it}$ : coeficiente beta de la cartera P.
- $\tilde{\varepsilon}_{pt} = \sum_{i=1}^{H \leq N} w_i^p \tilde{\varepsilon}_{it}$ : perturbación aleatoria de la cartera P.

El perfil de expresión de cualquier cartera está, de la misma forma que el perfil de expresión de cualquier gen, cualitativamente afectada por los tres componentes (alfa, beta y épsilon). Una muestra tumoral P se diferenciará de otra P' igualmente integrada por el mismo número (H) y tipo de genes que aquélla en el hecho de que a alguno de tales genes tendrá su expresión distinta y por tanto un “peso” o porcentaje de contribución al perfil de expresión de la muestra distinto. ( $w_i^P \neq w_i^{P'}$ ). En consecuencia, el perfil de expresión de dichas carteras diferirá.

El modelo de mercado no proporciona una relación funcional entre el perfil de expresión y el riesgo de la muestra sino entre el perfil de expresión de dicha muestra y de la muestra de control (cartera de mercado).

La aplicación del modelo de estimación mínimo cuadrático exige, como es sabido, un conjunto de hipótesis previas con relación a la perturbación aleatoria que, en el ámbito del modelo de mercado y para el caso del gen i, son las siguientes:

1. En cualquier período t los distintos factores que, en forma de perturbaciones aleatorias, afecten, en su caso, a la expresión de un gen tiene un efecto esperado conjunto nulo sobre la expresión esperada del mismo (los efectos positivos y los negativos se compensan mutuamente en su globalidad):

$$E[\tilde{\varepsilon}_{it}] = 0 \quad , \quad \forall t$$

2. En cualquier período  $t$  la varianza de cualquier perturbación aleatoria, que, en su caso, afecte a la expresión de un gen determinado es la misma (esto es, tiene un valor constante igual a  $\sigma_{\tilde{\varepsilon}_i}^2$ ):

$$\sigma_{\tilde{\varepsilon}_{it}}^2 = \sigma_{\tilde{\varepsilon}_i}^2, \quad \forall t$$

3. Cualquier perturbación aleatoria que, en su caso, afecte a la expresión de un gen determinado no dependen en modo ni período alguno de la evolución de la enfermedad (representada por la evolución de la cartera  $M$ ); por lo tanto la covarianza entre ambas variables es nula:

$$\sigma_{\tilde{\varepsilon}_{it}\tilde{R}_{Mt}}^2 = 0, \quad \forall t$$

4. Cualquier perturbación aleatoria que, en su caso, afecte a la expresión de un gen determinado a lo largo de un período de tiempo no influye ni está influida por cualquier otra que pueda afectar a dicho gen en otro período distinto:

$$\sigma_{\tilde{\varepsilon}_{it}\tilde{\varepsilon}_{it'}}^2 = 0, \quad \forall t \neq t'$$

5. Cualquier perturbación aleatoria que afecte a la expresión de un gen determinado sigue en cualquier período una distribución normal; por tanto, se tratará de una normal de media cero y varianza constante igual a  $\sigma_{\tilde{\varepsilon}_i}^2$  (hipótesis de normalidad):



$$\tilde{\varepsilon}_{it} \rightarrow N[0, \sigma_{\tilde{\varepsilon}_i}^2] \quad , \quad \forall t$$

Este modelo nos permitiría clasificar a los pacientes no sólo teniendo en cuenta los datos histopatológicos que clasifican al cáncer en cuatro estadios, sino atendiendo, además, a su perfil de expresión. Esto es de vital importancia porque podría arrojar alguna luz al hecho de que en ocasiones un paciente diagnosticado con un cáncer en estado avanzado responde mejor al tratamiento que otro paciente con el mismo cáncer pero con mejor pronóstico al encontrarse en un estadio más temprano. Del mismo modo podría facilitarnos información, atendiendo a los perfiles de expresión, de porqué un grupo de pacientes diagnosticados sus tumores en el mismo estadio, unos responden mejor al tratamiento que otros, incluso el por qué otros ni siquiera responden.

De verificarse lo anterior, que podemos clasificar mediante el perfil de expresión de la muestra tumoral de un paciente con cáncer, abriría la puerta a la posible elaboración de tratamientos más efectivos y personalizados para cada paciente en función de su perfil de expresión.

#### **2.2.4. Riesgo de una muestra (cartera)**

Todo lo dicho del riesgo de un gen es aplicable al riesgo de una cartera.

$$\sigma_{\tilde{R}_{pt}}^2 = \beta_p^2 \sigma_{\tilde{R}_{Mt}}^2 + \sigma_{\tilde{\varepsilon}_{pt}}^2$$

$$\sigma_{\tilde{\epsilon}_{pt}}^2 = \sum_{i=1}^{H \leq N} (w_i^p)^2 \sigma_{\tilde{\epsilon}_{it}}^2$$

$$\beta_p^2 = \left[ \sum_{i=1}^{H \leq N} w_i^p \beta_i \right]^2$$

Se observa que, al igual que sucede con un único gen, el riesgo total de una muestra (cartera) P ( $\sigma_{\tilde{R}_{pt}}^2$ ) podría descomponerse en riesgo sistemático ( $\beta_p^2 \sigma_{\tilde{R}_{Mt}}^2$ ) y riesgo específico ( $\sigma_{\tilde{\epsilon}_{pt}}^2$ ), si bien en el caso de la muestra la magnitud de ambos tipos de riesgo depende tanto de los genes escogidos, que en nuestro caso son los mismos para todas las muestras, como de la “proporción” o peso que éstos tengan en el perfil de expresión de su muestra ( $w_i^p$ ,  $i=1, \dots, H$ ).

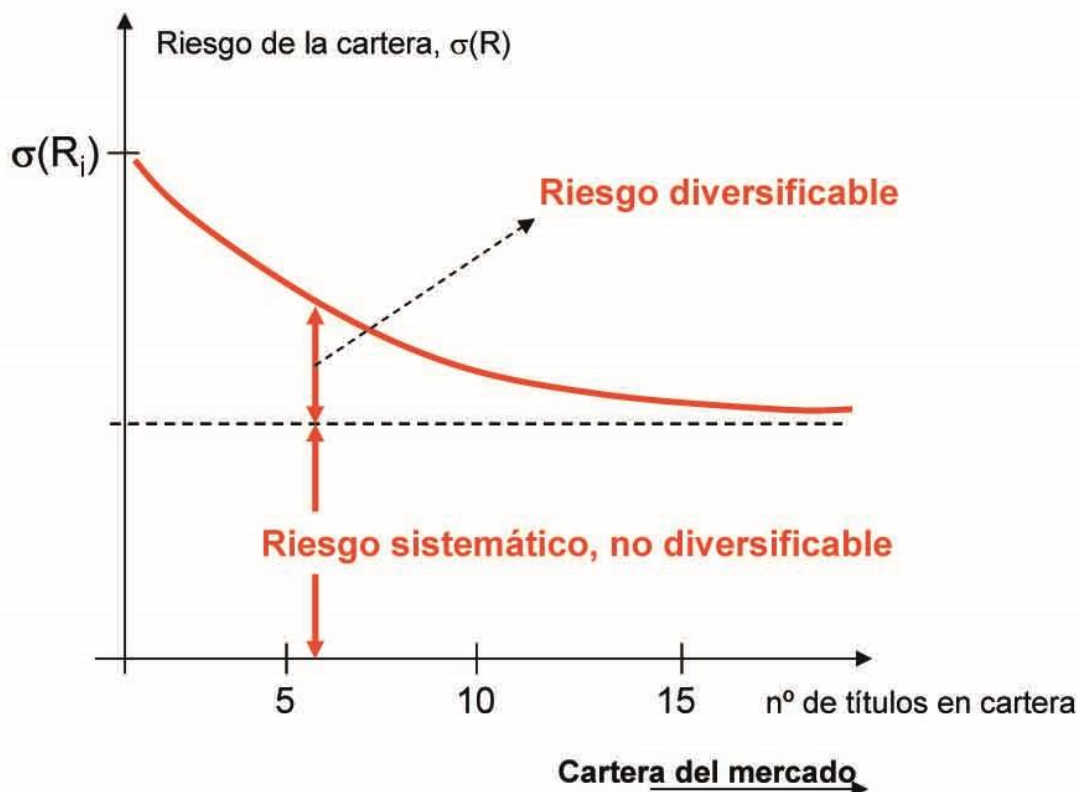


Figura 2.10.

Fuente: elaboración propia

En el eje horizontal se representa el número de genes que forman parte de la cartera, mientras que en el vertical se mide el riesgo de forma relativa frente a esa cartera ideal. La línea curva nos marca el riesgo total asociado con una cartera. Descontando del del valor total, el valor de mercado por la línea recta de puntos tendremos el riesgo específico. La línea de puntos representa así la parte del riesgo debida a la incertidumbre de la cartera de mercado.

## CAPÍTULO III. METODOLOGÍA

### 3.1. BASES DE DATOS Y JUSTIFICACIÓN DE LAS MUESTRAS

Los datos han sido cedidos por el Departamento de Oncología Molecular del Hospital Clínico de Madrid entre el año 2001 y el 2006. Los pacientes han sido seleccionados de manera que todos los estadios<sup>54</sup> del tumor estuvieran presentes en el estudio (24 en el estadio A, 26 en el estadio B, 19 en el estadio C y 19 en el estadio D), más o menos de manera homogénea.

A continuación se detalla el proceso seguido desde el momento en el que al paciente se le realiza la cirugía para la extracción del tumor hasta que los datos sobre la expresión de los genes son obtenidos en el Laboratorio e incluidos en una hoja de Excel.

La extirpación del tumor mediante cirugía, no se limita simplemente a la extracción de la masa tumoral; sino que para tratar de prevenir pequeñas micrometástasis<sup>55</sup> o células colindantes que se hayan escapado ya del tumor, se elimina una zona más amplia. Esto también es necesario para el análisis patológico del tumor; pues la clasificación por estadios<sup>56</sup> exige el análisis de los ganglios linfáticos<sup>57</sup> cercanos y distantes al tumor con el fin de determinar su grado de extensión y la posible metástasis. La zona de seguridad que se elimina

---

<sup>54</sup> Los estadios mencionados corresponden con la clasificación DUKES.

<sup>55</sup> Número pequeño de células cancerosas que se han diseminado desde el tumor primario a otras partes del cuerpo y que son muy pocas como para que una prueba de detección o diagnóstico las encuentre.

<sup>56</sup> Clasificación realizada al tumor en función de su tamaño, infiltración a ganglios y metástasis.

<sup>57</sup> Masa redondeada de tejido linfático rodeada por una cápsula de tejido conjuntivo. Los ganglios linfáticos filtran la linfa (líquido linfático), y almacenan los linfocitos (glóbulos blancos). Están ubicados a lo largo de los vasos linfáticos. También se llama glándula linfática.

en torno al tumor se llama zona de resección, y tal y como se ha comentado, cumple un doble propósito.

En la cirugía también se extrae una muestra sana del colon que debe estar a diez centímetros de la línea de resección. Estas dos muestras, la tumoral y la normal, son enviadas por el cirujano al patólogo para su posterior análisis.

En el Departamento de Patología, seccionan en distintas porciones ambas muestras, la tumoral y la del tejido sano. Algunas de estas porciones son utilizadas para establecer la clasificación histopatológica del tumor, las restantes se congelan y son enviadas al Banco de Tumores del Hospital Clínico.

El Laboratorio de Oncología Molecular del Hospital Clínico pide al Banco de Tumores las muestras para su posterior análisis. Una vez obtenidas, lo primero que debe realizarse es un estudio histopatológico de éstas con el fin de determinar qué porcentaje de tejido tumoral se encuentra en cada muestra. Ocurre, en ocasiones, que la masa tumoral contiene también, por ejemplo, tejido muscular. Hay que estar seguro de que esto no suceda; para ello se establece el criterio de que al menos el 80% de la muestra debe ser tejido tumoral. Las muestras que no superen esta condición no formarán parte del estudio.

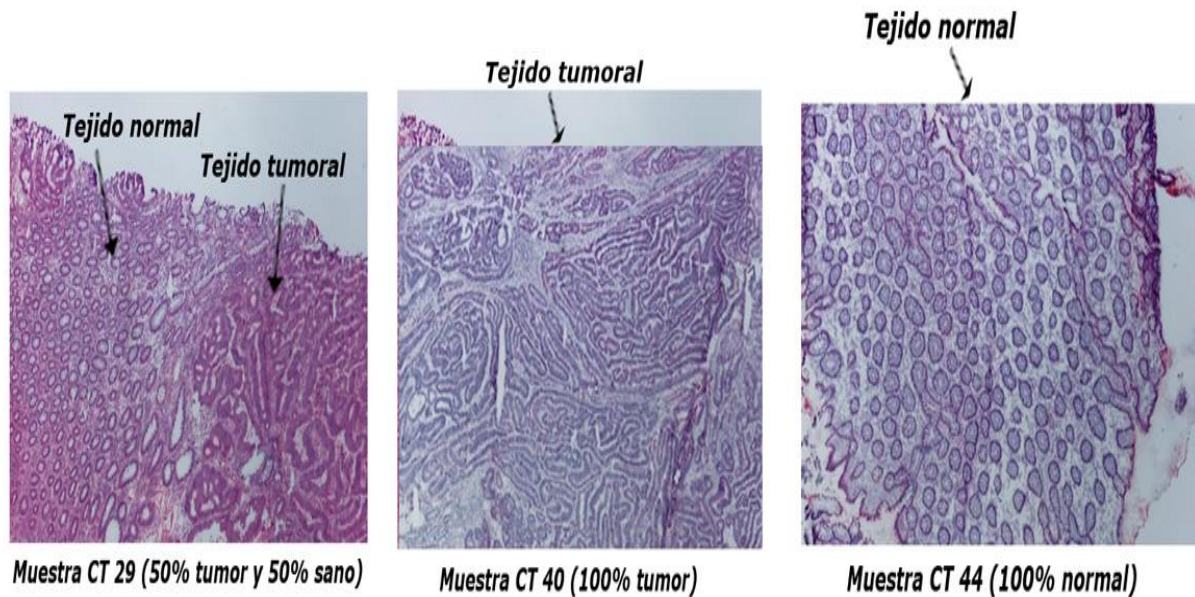


Imagen 3.1.

Fuente: Departamento de Oncología Molecular del Hospital Clínico de Madrid.

La imagen de la muestra CT29 de la figura 4.1., proviene de la muestra CT 29 (las letras CT indican que es una muestra tumoral). En ella se observa que el 50% del tejido corresponde a tejido sano y el otro 50% a tejido tumoral. En la imagen correspondiente a la muestra CT 40, el 100% del tejido es tumoral, mientras que en la última, el 100% del tejido es tejido sano o normal. Aplicando el criterio establecido anteriormente, las muestras CT 29 y CT 44 no formarían parte del estudio. Cuando esto sucede, y una muestra no supera el 80% de tejido tumoral; se pide nuevamente al Banco de Tumores que suministren, si es posible, otra porción del tumor congelado con el fin de no eliminar del estudio a ningún paciente. Uno de los problemas con los que se encuentran los laboratorios a la hora de realizar sus investigaciones es la dificultad existente para reunir muestras.

El siguiente paso es medir la purificación del ARN (del Inglés Ácido Ribonucleico) mediante RIN (Rna Integer Number, este algoritmo permite la clasificación de las muestras de RNA eucariótico total con base a un sistema numérico de 1 a 10, siendo 1 el valor para una muestra de RNA totalmente degradada y 10 el valor obtenido para un muestra intacta)

Se realiza un pool de ARN de tejido de colon normal que consiste en lo siguiente: se seleccionan 68 muestras al azar de tejido sano de los pacientes, una por paciente. Estas muestras como se dijo anteriormente deben guardar una distancia mínima de diez centímetros con la línea de resección. Se mezcla todo el ARN (RIN=8,5). Este pool de ARN se divide en 96 partes alícuotas, tantas como muestras son objeto del estudio.

La siguiente fase es la tinción de las muestras tumorales y el pool mediante el proceso conocido como transcripción inversa. El ADN (del inglés Ácido Desoxiribonucleico), mediante la transcripción pasa a formar ARN (proceso conocido como replicación). Se puede inducir el proceso inverso, esto es, pasar de ARN a ADN mediante una proteína, lo que nos permite a su vez, la tinción del ADN. Se utiliza la Cyanina 3 (que da un color verde) para la tinción de la muestra del pool y de Cyanina 5 (que da un color rojo) para la tinción de la muestra tumoral o de control.

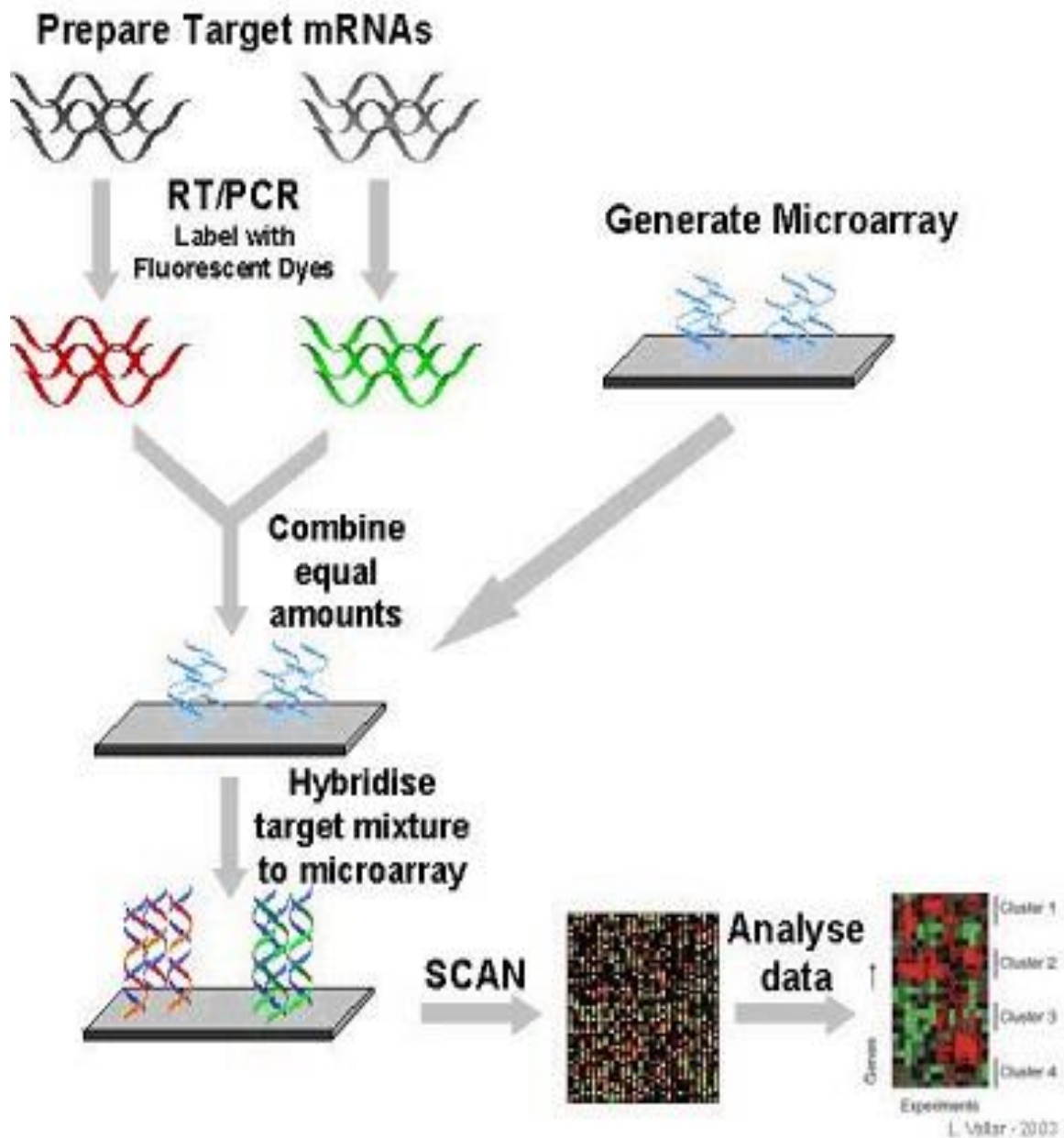


Imagen 3.2.

Fuente: [http://www.microarray.lu/images/overview\\_1.jpg](http://www.microarray.lu/images/overview_1.jpg)

Gracias a la nueva tecnología de microarrays, podemos obtener la expresión simultánea de todos los genes que forman el genoma humano. Una vez teñidas ambas muestras (verde tejido procedente del pool y rojo tejido tumoral o de control) se hibridan y su resultado queda expresado en los microarrays.





Imagen 3.3.

Fuente: <http://tetrad.ucsf.edu/bioreg/downloads/microarray.jpg>

Cada punto de los microarrays corresponde a un solo gen, siendo el resultado del cociente de la muestra del pool dividido entre la muestra tumoral o de control. Es decir, el resultado de cada punto o gen proviene de la división del valor del gen de la muestra del pool entre el valor del gen de la muestra tumoral o de control.

### 3.2. ANÁLISIS PRELIMINAR DE LAS MUESTRAS

Interpretaremos estos datos de dos maneras: gráfica y analíticamente

**Gráficamente**, si el color del punto es rojo significa que está sobre-expresado ese gen en esa muestra. Si la muestra seleccionada es tumoral nos indicará que ese gen está sobre-expresado en el tumor. Es decir, ese gen tiene una mayor actividad en la muestra tumoral que en la muestra del pool. Si por el contrario el punto es de color verde nos indica que el gen está inhibido en la muestra. Si la muestra es tumoral nos viene a decir que tiene menos actividad ese gen en el tumor que en el pool. Y por último, si el punto es de color amarillo implica que ese gen está igualmente expresado en ambas muestras, en la muestra tumoral o de control y en el pool.

#### ***Sección del microarray CT 44***

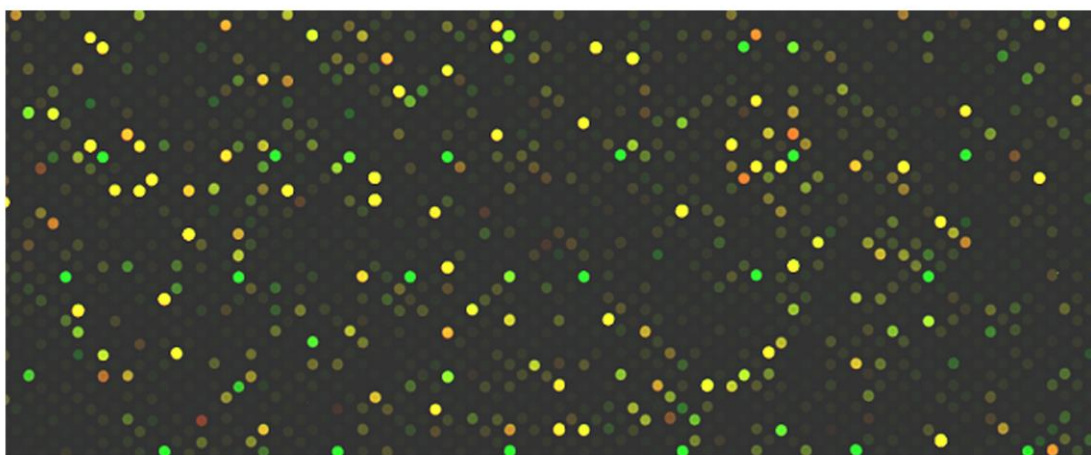
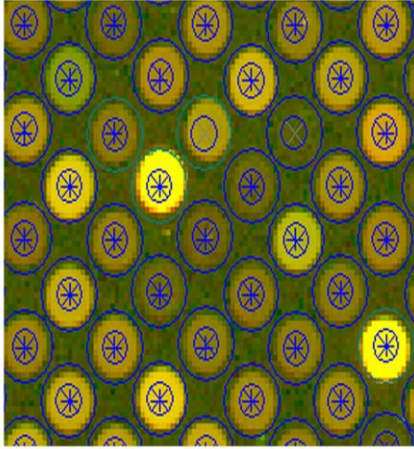


Imagen 3.4.

Fuente: Departamento de Oncología Molecular del Hospital Clínico de Madrid.

Esta fluorescencia se mide con un escáner de láser confocal y se cuantifica la fluorescencia de cada círculo y del fondo con el software Agilent Feature Extraction Software.



*El anillo interior indica la porción del círculo que utiliza en los cálculos*

*El anillo exterior indica la zona que se utiliza para sustraer el fondo*

Imagen 3.5.

Fuente: Departamento de Oncología Molecular del Hospital Clínico de Madrid.

**Analíticamente:** Los datos  $CT^{58}$  y  $CN^{59}$  están expresados como el cociente de logaritmos en base 2 de:

$M$  = valor numérico asignado al pool por el procedimiento del hospital. Es común para todos los datos.

$y_{ij}$  = valor numérico del gen  $i$  en la muestra  $j$ . Las muestras son 88 tumorales (CT) y 8 de control (CN).

---

<sup>58</sup> Identificativo para nombrar a las muestras tumorales

<sup>59</sup> Identificativo para nombrar a las muestras de control

Donde:

$i = 1, \dots, 14765$  número de genes motivo de estudio en todas las muestras

$j = 1, \dots, 96$  número de muestras. 89 tumorales (CT) y 7 de control (CN).

Definimos  $x_{ij} = \frac{Z}{y_{ij}}$

Todas las variables son mayores que cero.

Los datos los obtenemos mediante la siguiente expresión:

$$z_{ij} = \text{Log}_{(2)}(x_{ij}) = \text{Log}_{(2)}\left(\frac{M}{y_{ij}}\right)$$

Aunque matemáticamente  $z_{ij}$  puede tomar cualquier valor, biológicamente no tomará valores extremos.

Se espera que los valores de  $z_{ij}$  cercanos a cero correspondan a las muestras de control y a aquellos tumores que estén diferencialmente poco expresados.

Aquellos valores distintos de cero, tanto positivos como negativos, se espera que correspondan con los genes que están mas sobre-expresados (datos claramente negativos) o inhibidos (datos claramente positivos)

De las expresiones arriba mencionadas podemos deducir que: si el cociente es cero implicaría que en ambas muestras, tanto la tumoral y el pool por

un lado como la muestra de control y el pool por otro, sus numeradores y denominadores deberán ser iguales.

Si el cociente es igual a uno se verifica que:

$$\text{Log}_{(2)}(1) = 0.$$

Lo que implicaría que ese gen se expresa por igual tanto en la muestra tumoral (o de control) como en el pool. En otras palabras, la expresión de ese gen no varía en la muestra tumoral, o de control, con respecto al pool.

Cuanto más se aleje la expresión de un gen de cero más variación habrá en la expresión de ese gen en la muestra tumoral o de control con respecto a la muestra del pool (muestra sana).

Los datos por tanto pueden ser positivos y negativos:

- Un valor negativo nos indica que el numerador es menor que el denominador. Dicho de otro modo; que el valor del gen en el tumor tiene un valor mayor (está sobre-expresado) con relación al valor de ese gen en el pool.
- Un valor positivo nos indica que el numerador es mayor que el denominador. Dicho de otro modo; el valor de ese gen en el tumor tiene un valor menor (está inhibido) con relación al valor de ese gen en el pool.

El objetivo del presente estudio es validar si el parámetro beta, utilizado en la medición del riesgo en el análisis bursátil, en el modelo de regresión lineal puede ser utilizado para realizar una selección de genes que presenten un patrón de expresión diferencialmente expresado entre las muestras sanas y las tumorales. Actualmente, la investigación básica en el campo de la oncología está tratando de aportar medidas y técnicas válidas que sean capaces de determinar qué conjunto de genes se encuentra diferencialmente expresado en un conjunto de pacientes que presentan un tumor para así poder ofrecer tratamientos personalizados al paciente que sean a la vez más eficaces y con menos efectos secundarios que los actualmente existentes.

### 3.3. HIPÓTESIS DEL ESTUDIO

Se tomará como medida la diferencia existente entre la muestra a analizar y el pool como indicador de inestabilidad. Asumiremos como hipótesis que cuánto más pequeña sea esta distancia, en valor absoluto, potencialmente menos peligrosa será la muestra.

En un primer momento interesa determinar qué genes son los que tienen mayor variabilidad, independientemente del signo de ésta, pues se pretenden seleccionar aquellos que denoten mayor variabilidad. Posteriormente habrá que determinar, pues su sentido biológico dependerá de ello, si dichos genes están sobre-expresados o inhibidos (si su signo es positivo o negativo). Cuando la ratio entre muestra analizada y muestra del pool sea cero, como ya se ha explicado, indica que ese gen no varía, es decir, tiene la misma expresión en ambas muestras.

Son interesantes para este estudio aquellas  $z_{ij}$  que estén más distantes de cero, con independencia del signo, se toma el valor absoluto de dichas variables  $|z_{ij}|$ ; que serán las variables de partida para realizar el análisis.

Evidentemente, cada  $z_{ij}$  representa un solo gen en cada una de las muestras, y si realizamos el  $\sum_{i=1}^{14764} z_{ij} = d_j$  tendremos el peso total de todas las  $z_{ij}$  de la muestra j,

Cabe esperar que el valor de  $d_j$  sea grande en las muestras tumorales y los más pequeños en las muestras de control.

Como vamos a emplear la metodología habitualmente usada en el análisis bursátil para el cálculo de las rentabilidades de los diferentes títulos

cotizados en los mercados financieros, definimos la variable  $r_{ij} = \frac{z_{ij}}{\bar{z}_{ic}} - 1$ ,

donde:

- $\bar{z}_{ic}$  es la media del gen i de los controles, de tal forma que  $r_{ij}$  representa una “rentabilidad”.
- Si el valor de  $z_{ij} > \bar{z}_{ic}$  aporta más a la variabilidad, en tanto por uno, el gen i a la muestra j.
- Si el valor de  $z_{ij} < \bar{z}_{ic}$  aporta poco a la variabilidad, en tanto por uno, el gen i a la muestra j.

De la misma manera definimos  $R_j = \frac{d_j}{\bar{d}_c} - 1$

Donde:

- $\bar{d}_c$  representa la media para las muestras de control. En este caso  $R_j$  representa la variabilidad, en tanto por uno, de la muestra j respecto a las muestras de control.



Ordenadas las muestras en función del valor de  $d_j$ , de menor a mayor, observamos que, efectivamente, las muestras de control, sin excepción, son las que menos variabilidad presentan.

### 3.4. TÉCNICAS EMPLEADAS

#### 3.4.1. Modelo de Regresión Simple

El modelo de regresión lineal simple es una clase de modelo econométrico con las siguientes características: lineal, uniecuacional, estático, con una variable endógena, una variable exógena, un término de error y dos parámetros.

La formulación teórica es:

$$Y = \beta_1 + \beta_2 X + u$$

Donde Y representa a la variable endógena, X a la variable exógena, u es la variable aleatoria que representa al término de error y  $\beta_1$  y  $\beta_2$  son los parámetros.

Si consideramos los “N” elementos muestrales que se van a observar:

$$Y_1 = \beta_1 + \beta_2 X + u_1$$

$$Y_2 = \beta_1 + \beta_2 X + u_2$$

$$Y_i = \beta_1 + \beta_2 X + u_N$$

Siendo:

- $Y_i$ : observaciones muestrales de la variable endógena Y, variable cuyo comportamiento pretende explicar el modelo.
- $X_i$ : observaciones muestrales de la variable exógena X, variable elegida como explicativa e influyente sobre la variable Y.

- $u_i$ : elementos muestrales latentes o no observables de la variable  $u$ , se denominan también perturbaciones aleatorias o términos de error. Representan al conjunto de causas explicativas de la variable  $Y$ , desconocidas o incontrolables y por tanto no incluidas en la estructura del modelo.
- $\beta_1, \beta_2$ : son los parámetros estructurales del modelo que permanecen invariantes respecto a la variación muestral.  $\beta_1$  es el término independiente u ordenada en el origen,  $\beta_2$  es la pendiente o coeficiente de regresión que es el parámetro que acompaña y pondera a la variable exógena  $X$ .

Recordamos en este punto que las hipótesis que el modelo de regresión lineal simple son:

1. La relación es estocástica (aleatoria): el término de error  $\varepsilon_i$  recoge la componente aleatoria de  $y_i$  que el modelo no puede explicar.  $\varepsilon_i$  es no observable.
2. Ausencia de error de especificación. Aparecen en el modelo todas las variables relevantes para explicar el comportamiento de  $Y$ . Esta hipótesis está directamente relacionada con la investigación inicial ya que es el investigador aplicado quien conoce, a partir de la teoría, que variables pueden ser relevantes para explicar la respuesta. Esta hipótesis es necesaria para que el término de perturbación aleatoria sea error puro con

media nula.

3. Linealidad de la relación.  $E(y) = X\beta$  Las medidas de las distribuciones de Y condicionadas a cada valor de X se encuentran sobre una línea (en el caso más simple).
4. Esperanza matemática nula del término de perturbación: La especificación correcta del modelo hace que no se introduzca ninguna componente sistemática en los errores al compensarse, en promedio, los positivos y negativos. Esta hipótesis es consecuencia directa de la anterior.
5. Homocedasticidad : Varianza constante de los errores:  $\text{var}(\varepsilon_i) = \sigma^2$ , para todo i.
6. No autocorrelación: Ausencia de covarianza (o correlación) entre los errores:  $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$  si  $i \neq j$ .
7. Variables explicativas deterministas o no aleatorias. Variables controladas por el investigador y medidas sin error. De esta forma el modelo lineal general está especialmente indicado en el análisis de experimentos diseñados en los que se controlan las condiciones experimentales. Esta hipótesis se puede relajar suponiendo que las variables regresoras son independientes del error aunque no sean constantes. En la mayor parte de las aplicaciones del modelo lineal las variables regresoras son aleatorias.

8. No multicolinealidad: Es decir las variables explicativas no son linealmente dependientes (ninguna de ellas puede obtenerse como combinación lineal de las demás). El problema será estudiado posteriormente con más detalle.
9. Constancia de los parámetros: Debemos admitir una única estructura válida para el periodo de observación y el horizonte de predicción.

### 3.4.2. Criterio de Mínimos Cuadrados Ordinarios

A continuación se expone el proceso de obtención de los estimadores de MCO. El objetivo es minimizar la suma de los cuadrados de los residuos (S). Para ello, en primer lugar expresamos S como una función de los estimadores.

Por lo tanto:

$$\text{Min}_{\hat{\beta}_1, \hat{\beta}_2} S = \text{Min}_{\hat{\beta}_1, \hat{\beta}_2} \sum_{i=1}^n \hat{u}_i^2 = \text{Min}_{\hat{\beta}_1, \hat{\beta}_2} \sum_{i=1}^n (Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_i)^2 \quad (3-1)$$

Para minimizar S, derivamos parcialmente con respecto a  $\hat{\beta}_1$  y  $\hat{\beta}_2$ :

$$\frac{\partial S}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_i)$$

$$\frac{\partial S}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_i) X_i$$

Los estimadores de MCO se obtienen igualando las anteriores derivadas a cero:

$$\sum_{i=1}^n (Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_i) = 0 \quad (3-2)$$

$$\sum_{i=1}^n (Y_i - \hat{\beta}_1 + \hat{\beta}_2 X_i) X_i = 0 \quad (3-3)$$

Las ecuaciones (3-1 y 3-2) se denominan ecuaciones normales o condiciones de primer orden de MCO.

Operando con las ecuaciones normales, (obviaremos el proceso matemático para llegar a determinar los estimadores) se obtiene  $\hat{\beta}_2$ :

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i x_i - \bar{y} \sum_{i=1}^n x_i}{\sum_{i=1}^n x_i^2 - \bar{x} \sum_{i=1}^n x_i}$$

O también:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Siendo  $\bar{y}, \hat{\beta}_1$ :

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 X$$

En nuestro caso, el estudio pretende realizar un ajuste lineal entre la variable  $r_i$  que son las variabilidades definidas anteriormente en el gen i en las 96 muestras y R que son las variabilidades totales de cada una de las 96 muestras.

Utilizando el modelo de regresión lineal simple de mínimos cuadrados ordinarios (M.C.O.), la  $r_i$  es la variable respuesta (variable dependiente) y R la variable explicativa (variable independiente). Dicha función queda expresada de la siguiente manera:

$$r_i = \beta_i R + \alpha_1$$

### **3.3. DESCRIPCIÓN, ANÁLISIS Y RESULTADOS**

#### **3.3.1. Descripción**

Los datos iniciales, normalizados y entregados por el laboratorio, son mostrados en una hoja de Excel llamada "DATOS". Las columnas corresponden a las muestras tumorales y normales, una muestra por columna, mientras que las filas corresponden a los genes, una fila por cada gen.

La muestra tumoral es, como se ha explicado anteriormente, el cociente entre la muestra del pool y la muestra procedente del tumor. Por otro lado se han realizado muestras de control para asegurar la calidad del estudio. Dichas muestras son el cociente resultante entre el pool y las muestras de control (recordar que el pool había sido formado con las muestras de tejido sano de colon de 68 pacientes al azar) y una muestra de tejido sano de un paciente. Las muestras de control son muestras de tejido sano que han sido utilizadas en la elaboración del pool. El proceso que se ha seguido es el mismo que para las muestras de tumorales. Se tiñen con cyanina 3 (color verde) las muestras de tejido normal y con cyanina 5 (color rojo) las muestras del pool.

Para identificar las muestras de control de las tumorales se ha nombrado a las muestras tumorales con el prefijo CT, seguido del número de la muestra (por ejemplo la muestra CT 44 corresponde con la muestra tumoral 44). Las muestras normales van precedidas del prefijo CN seguido del número de la



muestra (por ejemplo la muestra CN 27 corresponde a la muestra normal 27).

De las 96 muestras, 7 son normales y 89 tumorales.

**Columna identificadores de genes** (points to column B)

**Muestra tumoral CT 20** (points to column F)

**Fila de indicadores de muestras** (points to row 2)

**Fila del gen NM\_178493** (points to row 7)

**Expresión del gen NM\_000064 en la muestra CT27** (points to cell D10)

	B	C	D	E	F	G	H	I
2	Unique id	CT 16	CT 27	CT 31	CT 20	CT 28	CT 30	CT 36
3	NM_194272	-1.20768297	-1.6404227	-1.72779357	-1.79193628	-1.69913971	-2.17404962	-0.14555988
4	NM_004062	-0.96689433	-1.0608288	-0.88806778	-1.10523331	-0.72629058	-0.90001947	-0.74723399
5	NM_002615	-0.14605165	-1.23968148	0.08376773	-1.10700166	-1.00123525	-0.9385764	2.31302261
6	NM_003726	0.14283906	-0.17363966	0.62029636	0.38083225	0.72730881	-1.75384259	0.19457261
7	NM_178493	0.52026832	0.55069405	0.75904423	-0.29431161	0.21255769	2.05335665	-0.07226272
8	NM_002091	3.24920583	0.0134244	1.31962848	0.70652562	0.14833754	2.30398965	2.31058311
9	NM_003012	-2.47517395	-3.1222415	-2.64318776	-2.86929274	-3.0469346	-2.9505169	2.20614338
10	NM_000064	-0.88549894	-3.05912662	-1.20998883	-2.6570797	-2.51888561	-1.55599356	2.13611603
11	NM_144947	1.74229825	-0.68513161	-1.03869872	-0.62309867	3.08039927	1.34485948	-0.78186738
12	NM_152358	2.62514043	1.4675113	1.60514801	4.1479001	1.54552277	1.1315074	2.416084
13	NM_002842	-1.11111111	-1.11111111	-1.11111111	-2.03880477	-2.28850937	-1.95340884	-2.77557898
14	NM_001828	-1.28000000	-1.28000000	-1.87753737	-1.38253045	-1.83976722	-1.39230192	-2.09768844
15	NM_003500	0.26739531	0.64465248	0.16014498	0.16333811	-1.26289964	0.32353598	-0.27973726
16	NM_006887	-0.53123598	0.14586671	-0.61569566	-2.33182001	-2.69385171	-0.16363753	-0.07340594

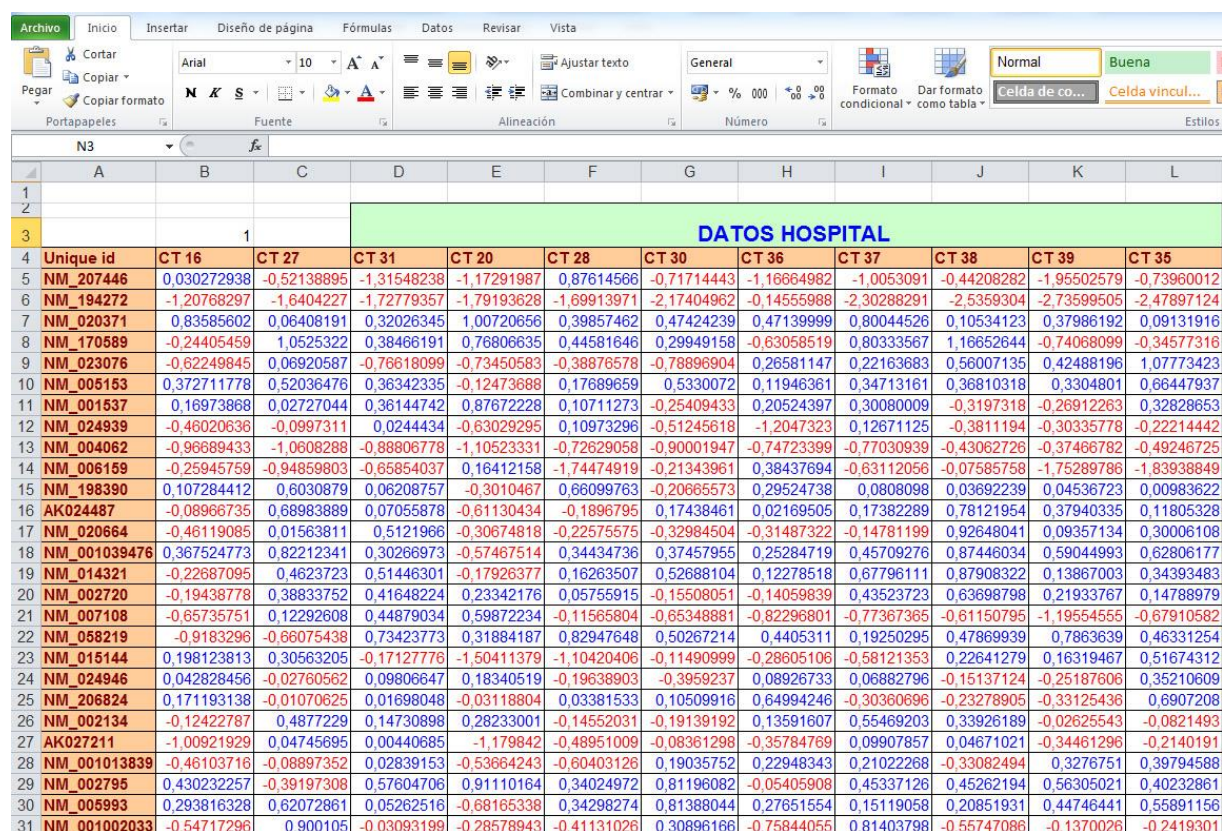
Imagen 3.6. Imagen de los datos correspondientes a muestras tumorales

Fuente: elaboración propia.

La tabla inicial de la hoja datos de Excel de este estudio consta de 97 columnas (la primera columna corresponde al identificador de los genes y las

otras 96 restantes a las muestras), y 14765 filas (la primera fila son los identificadores de las muestras y las otras 14764 corresponden a los genes).

Los datos facilitados por el Hospital Clínico están recogidos en la hoja de cálculo “Hospital Datos” en la tabla “DATOS HOSPITAL”.



Unique id	CT 16	CT 27	CT 31	CT 20	CT 28	CT 30	CT 36	CT 37	CT 38	CT 39	CT 35
NM_207446	0.030272938	-0.52138895	-1.31548238	-1.17291987	0.87614566	-0.71714443	-1.16664982	-1.0053091	-0.44208282	-1.95502579	-0.73960012
NM_194272	-1.20768297	-1.6404227	-1.72779357	-1.79193628	-1.69913971	-2.17404962	-0.14555988	-2.30288291	-2.5359304	-2.73599505	-2.47897124
NM_020371	0.83585602	0.06408191	0.32026345	1.00720656	0.39857462	0.47424239	0.47139999	0.80044526	0.10534123	0.37986192	0.09131916
NM_170589	-0.24405459	1.0525322	0.38466191	0.76806635	0.44581646	0.29949158	-0.63058519	0.80333567	1.16652644	-0.74068099	-0.34577316
NM_023076	-0.62249845	0.06920587	-0.76618099	-0.73450583	-0.38876578	-0.78896904	0.26581147	0.22163683	0.56007135	0.42488196	1.07773423
NM_005153	0.372711778	0.52036476	0.36342335	-0.12473688	0.17689659	0.5330072	0.11946361	0.34713161	0.36810318	0.3304801	0.66447937
NM_001537	0.16973868	0.02727044	0.36144742	0.87672228	0.10711273	-0.25409433	0.20524397	0.30080009	-0.3197318	-0.26912263	0.32828653
NM_024939	-0.46020636	-0.0997311	0.0244434	-0.63029295	0.10973296	-0.51245618	-1.2047323	0.12671125	-0.3811194	-0.30335778	-0.22214442
NM_004062	-0.96689433	-1.0608288	-0.88806778	-1.10523331	-0.72629058	-0.90001947	-0.74723399	-0.77030939	-0.43062726	-0.37466782	-0.49246725
NM_006159	-0.25945759	-0.94859803	-0.65854037	0.16412158	-1.74474919	-0.21343961	0.38437694	-0.63112056	-0.07585758	-1.75289786	-1.83938849
NM_198390	0.107284412	0.6030879	0.06208757	-0.3010467	0.66099763	-0.20665573	0.29524738	0.0808098	0.03692239	0.04536723	0.00983622
AK024487	-0.08966735	0.68983889	0.07055878	-0.61130434	-0.1896795	0.17438461	0.02169505	0.17382289	0.78121954	0.37940335	0.11805328
NM_020664	-0.46119085	0.01563811	0.5121966	-0.30674818	-0.22575575	-0.32984504	-0.31487322	-0.14781199	0.92648041	0.09357134	0.30006108
NM_001039476	0.367524773	0.82212341	0.30266973	-0.57467514	0.34434736	0.37457955	0.25284719	0.45709276	0.87446034	0.59044993	0.62806177
NM_014321	-0.22687095	0.4623723	0.51446301	-0.17926377	0.16263507	0.52688104	0.12278518	0.67796111	0.87908322	0.13867003	0.34393483
NM_002720	-0.19438778	0.38833752	0.41648224	0.23342176	0.05755915	-0.15508051	-0.14059839	0.43523723	0.63698798	0.21933767	0.14788979
NM_007108	-0.65735751	0.12292608	0.44879034	0.59872234	-0.11565804	-0.65348881	-0.82296801	-0.77367365	-0.61150795	-1.19554555	-0.67910582
NM_058219	-0.9183296	-0.66075438	0.73423773	0.31884187	0.82947648	0.50267214	0.4405311	0.19250295	0.47869939	0.7863639	0.46331254
NM_015144	0.198123813	0.30563205	-0.17127776	-1.50411379	-1.10420406	-0.11490999	-0.28605106	-0.58121353	0.22641279	0.16319467	0.51674312
NM_024946	0.042828456	-0.02760562	0.09806647	0.18340519	-0.19638903	-0.3959237	0.08926733	0.06882796	-0.15137124	-0.25187606	0.35210609
NM_206824	0.171193138	-0.01070625	0.01698048	-0.03118804	0.03381533	0.10509916	0.64994246	-0.30360696	-0.23278905	-0.33125436	0.6907208
NM_002134	-0.12422787	0.4877229	0.14730898	0.28233001	-0.14552031	-0.19139192	0.13591607	0.55469203	0.33926189	-0.02625543	-0.0821493
AK027211	-1.00921929	0.04745695	0.00440685	-1.179842	-0.48951009	-0.08361298	-0.35784769	0.09907857	0.04671021	-0.34461296	-0.2140191
NM_001013839	-0.46103716	-0.08897352	0.02839153	-0.53664243	-0.60403126	0.19035752	0.22948343	0.21022268	-0.33082494	0.3276751	0.39794588
NM_002795	0.430232257	-0.39197308	0.57604706	0.91110164	0.34024972	0.81196082	-0.05405908	0.45337126	0.45262194	0.56305021	0.40232861
NM_005993	0.293816328	0.62072861	0.05262516	-0.68165338	0.34298274	0.81388044	0.27651554	0.15119058	0.20851931	0.44746441	0.55891156
NM_001002033	-0.54717296	0.900105	-0.03093199	-0.28578943	-0.41131026	0.30896166	-0.75844055	0.81403798	-0.55747086	-0.1370026	-0.2419301

Imagen 3.7. Datos Cedidos por el Hospital Clínico San Carlos de Madrid.

Fuente: elaboración propia.





control) sería 1; por tanto  $\log(2) = 0$ . En cambio si la expresión del gen se modificara de manera significativa en la muestra tumoral, quedaría reflejado en el dato del gen en el tumor, se alejaría del valor cero. Esto podría sugerir que la expresión de ese gen pudiera ser importante en la evolución del tumor.

Se asumirá como hipótesis de partida que cuánto más pequeña sea esta distancia, en valor absoluto, potencialmente será menos peligrosa dicha muestra. Las muestras de control, formadas por tejido sano, deberían tener muy poca variabilidad en relación con las muestras tumorales.

La variabilidad de una muestra puede quedar explicada por tres vías:

- Que muchos genes modifique su expresión de una manera no muy significativa individualmente.
- Que pocos genes modifiquen su expresión de una manera significativa.
- Que muchos genes modifiquen su expresión de manera significativa.

Por tanto, y con carácter general, diremos que cuánto menos variabilidad tenga una muestra menos malignidad presentará el tumor.

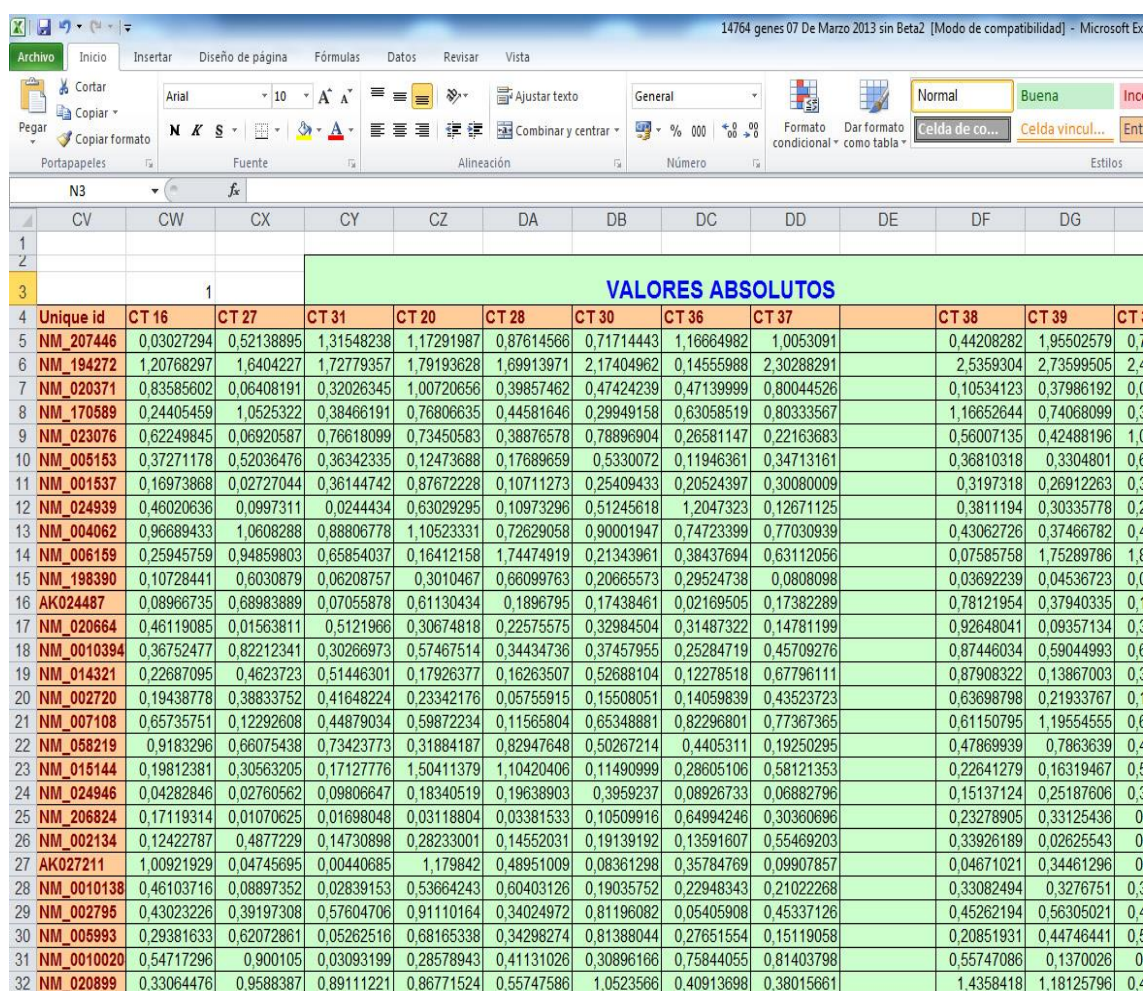
Para calcular la variabilidad de cada una de las 96 muestras se realizan los siguientes pasos:

1. **Calcular el valor absoluto.** Lo primero a realizar es obtener el valor absoluto de todos los genes de cada una de las muestras, tanto de las



tumorales como de las de control. Se prescinde, por ahora, del signo; pues lo interesante es identificar aquellos genes que expresan una mayor variabilidad. Por tanto, es importante identificar la variación de los genes independientemente que ello implique que estén sobre-expresados o inhibidos.

Los cálculos están realizados en la hoja de cálculo “Datos Hospital” en la tabla “VALORES ABSOLUTOS”



	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG
1												
2												
3		1										
4	Unique id	CT 16	CT 27	CT 31	CT 20	CT 28	CT 30	CT 36	CT 37		CT 38	CT 39
5	NM_207446	0.03027294	0.52138895	1.31548238	1.17291987	0.87614566	0.71714443	1.16664982	1.0053091		0.44208282	1.95502579
6	NM_194272	1.20768297	1.6404227	1.72779357	1.79193628	1.69913971	2.17404962	0.14555988	2.30288291		2.5359304	2.73599505
7	NM_020371	0.83585602	0.06408191	0.32026345	1.00720656	0.39857462	0.47424239	0.47139999	0.80044526		0.10534123	0.37986192
8	NM_170589	0.24405459	1.0525322	0.38466191	0.76806635	0.44581646	0.29949158	0.63058519	0.80333567		1.16652644	0.74068099
9	NM_023076	0.62249845	0.06920587	0.76618099	0.73450583	0.38876578	0.78896904	0.26581147	0.22163683		0.56007135	0.42488196
10	NM_005153	0.37271178	0.52036476	0.36342335	0.12473688	0.17689659	0.5330072	0.11946361	0.34713161		0.36810318	0.3304801
11	NM_001537	0.16973868	0.02727044	0.36144742	0.87672228	0.10711273	0.25409433	0.20524397	0.30080009		0.3197318	0.26912263
12	NM_024939	0.46020636	0.0997311	0.0244434	0.63029295	0.10973296	0.51245618	1.2047323	0.12671125		0.3811194	0.30335778
13	NM_004062	0.96689433	1.0608288	0.88806778	1.10523331	0.72629058	0.90001947	0.74723399	0.77030939		0.43062726	0.37466782
14	NM_006159	0.25945759	0.94859803	0.65854037	0.16412158	1.74474919	0.21343961	0.38437694	0.63112056		0.07585758	1.75289786
15	NM_198390	0.10728441	0.6030879	0.06208757	0.3010467	0.66099763	0.20665573	0.29524738	0.0808098		0.03692239	0.04536723
16	AK024487	0.08966735	0.68983889	0.07055878	0.61130434	0.1896795	0.17438461	0.02169505	0.17382289		0.78121954	0.37940335
17	NM_020664	0.46119085	0.01563811	0.5121966	0.30674818	0.22575575	0.32984504	0.31487322	0.14781199		0.92648041	0.09357134
18	NM_0010394	0.36752477	0.82212341	0.30266973	0.57467514	0.34434736	0.37457955	0.25284719	0.45709276		0.87446034	0.59044993
19	NM_014321	0.22687095	0.4623723	0.51446301	0.17926377	0.16263507	0.52688104	0.12278518	0.67796111		0.87908322	0.13867003
20	NM_002720	0.19438778	0.38833752	0.41648224	0.23342176	0.05755915	0.15508051	0.14059839	0.43523723		0.63698798	0.21933767
21	NM_007108	0.65735751	0.12292608	0.44879034	0.59872234	0.11565804	0.65348881	0.82296801	0.77367365		0.61150795	1.19554555
22	NM_058219	0.9183296	0.66075438	0.73423773	0.31884187	0.82947648	0.50267214	0.4405311	0.19250295		0.47869939	0.7863639
23	NM_015144	0.19812381	0.30563205	0.17127776	1.50411379	1.10420406	0.11490999	0.28605106	0.58121353		0.22641279	0.16319467
24	NM_024946	0.04282846	0.02760562	0.09806647	0.18340519	0.19638903	0.3959237	0.08926733	0.06882796		0.15137124	0.25187606
25	NM_206824	0.17119314	0.01070625	0.01698048	0.03118804	0.03381533	0.10509916	0.64994246	0.30360696		0.23278905	0.33125436
26	NM_002134	0.12422787	0.48772229	0.14730898	0.28233001	0.14552031	0.19139192	0.13591607	0.55469203		0.33926189	0.02625543
27	AK027211	1.00921929	0.04745695	0.00440685	1.179842	0.48951009	0.08361298	0.35784769	0.09907857		0.04671021	0.34461296
28	NM_0010138	0.46103716	0.08897352	0.02839153	0.53664243	0.60403126	0.19035752	0.22948343	0.21022268		0.33082494	0.3276751
29	NM_002795	0.43023226	0.39197308	0.57604706	0.91110164	0.34024972	0.81196082	0.05405908	0.45337126		0.45262194	0.56305021
30	NM_005993	0.29381633	0.62072861	0.05262516	0.68165338	0.34298274	0.81388044	0.27651554	0.15119058		0.20851931	0.44746441
31	NM_0010020	0.54717296	0.900105	0.03093199	0.28578943	0.41131026	0.30896166	0.75844055	0.81403798		0.55747086	0.1370026
32	NM_020899	0.33064476	0.9588387	0.89111221	0.86771524	0.55747586	1.0523566	0.40913698	0.38015661		1.4358418	1.18125796

Imagen 3.9.Imagen de la tabla “VALORES ABSOLUTOS”

Fuente: elaboración propia.

En esta misma tabla se calcula la variabilidad de cada muestra sumando todas las variabilidades de los genes de una muestra. Estos datos están recogidos en la última fila de la tabla bajo el nombre “SUMATORIO”.

Al ordenar las muestras en función de su variabilidad, las muestras de control están las primeras, pues son las que menos varían. Esto corrobora nuestro supuesto de partida: cuanto más cercanas a cero estén las muestras menos “malignas” son. El que la variabilidad de las muestras de control no sean cero atiende a varias razones:

- La expresión de cada gen está influida, en cada momento, por varios factores como pueden ser la edad, sexo, etc.
- Los genes no tiene un valor de expresión único, sino que; en su estado normal oscilan dentro de un rango. Lo mismo que sucede, por ejemplo, con los valores del colesterol. Si una persona, en condiciones normales, tiene unos niveles de colesterol entre 160 y 200 mg/dl se dice que su nivel de colesterol es bueno. Si dicho nivel se sitúa entre 200 y 240 mg/dl su nivel de colesterol es Normal-Alto. Por último, si su nivel de colesterol supera los 240 mg/dl tiene un colesterol alto.

Las muestras ordenadas en función de su variabilidad están recogidas en la tabla “VALORES ABSOLUTOS ORDENADOS SEGÚN MERCADO”, perteneciente a la hoja de cálculo “Datos”.



Observando los datos nos se percibe que la diferencia entre la variabilidad de las muestras de control y la variabilidad de las muestras tumorales es considerable. Mientras que la variabilidad promedio de las muestras de control es de 3.783,0022; la variabilidad promedio de las muestras tumorales es de 8.622,1186. El promedio de las muestras tumorales es 2,2791 veces más que el promedio de las muestras de control.

	Unique id	CN2	CN8	CN54	CN42	CN83	CN85	CN78	CT 116	CT 89	CT 81	CT 3	CT 98	CT 29
14775														
14776														
14777														
14778														
14779														
14780	NM 207446	0.52269149	0.05587619	0.18602891	0.13957353	0.46463871	0.49131879	0.01725937	0.08070274	0.90519869	0.10255121	0.94278413	0.96401376	0.25989591
14781	NM 194272	0.33954856	1.02105045	0.27491599	0.87339729	0.1910193	0.21439877	0.87523091	0.44712973	0.9531306	1.57195449	1.65691817	1.0774281	1.8367121
14782	NM 020371	0.0390427	0.10202263	0.16701977	0.20236959	0.03868085	0.10903247	0.2561394	0.23025459	1.00977874	0.57112581	0.50755405	0.07138396	0.86040981
14783	NM 170589	0.14258635	0.59138739	0.14891638	0.26391885	0.1110931	0.31038031	0.1322366	0.25320739	0.16632891	0.64378858	0.52530146	1.70709205	0.85083841
14784	NM 023076	0.25266778	0.43255889	0.03632422	0.09380069	0.07228161	0.06056086	0.13881926	0.95030689	0.02519123	0.16288669	0.18457332	0.20474191	0.05076761
14785	NM 005153	0.02680916	0.00214372	0.174887	0.23738235	0.0005884	0.18020925	0.33980656	0.29632854	0.23964806	0.00910497	0.27173144	0.39955434	0.25492531
14786	NM 001537	0.00263172	0.02287082	0.0150416	0.07655582	0.05240978	0.17622726	0.05763962	0.18196532	0.7021184	0.10997371	0.41818136	0.29464284	0.04228771
14787	NM 024939	0.00705328	0.16106959	0.02187005	0.02263602	0.18632771	0.0308178	0.35179225	0.06298421	0.80876243	0.22010003	0.18816398	0.24163516	0.19327711
14788	NM 004062	0.12972197	0.10016637	0.08193757	0.16109194	0.035396	0.35980186	0.22122291	0.20806281	0.40014601	0.14042085	0.12308189	0.07866125	0.13267711
14789	NM 006159	0.41255161	0.44215357	0.33286658	0.70114595	0.09989033	0.31206182	0.26624885	0.87101138	0.08448986	0.59266919	0.62227011	1.14410698	1.29373531
14790	NM 198390	0.31650195	0.14335504	0.20283952	0.31108066	0.08525592	0.12889239	0.0698864	0.21779466	0.02504228	0.19398424	0.1987984	0.05357637	0.00966431
14791	AK024487	0.28438002	0.01200606	0.16780619	0.23574087	0.28322178	0.29098731	0.01297397	0.62959784	0.16213313	0.31410593	0.13714547	0.02922689	0.35053411
14792	NM 020664	0.52226162	0.39818501	0.88059682	0.34470767	0.14905769	0.26069275	0.11964962	0.09615921	0.19027457	0.27981326	0.03374527	0.5049594	0.22140151
14793	NM 001039476	0.25871184	0.3066082	0.08951968	0.27923825	0.19114614	0.10649594	0.15446772	0.50531137	0.46696138	0.5267576	0.50034356	0.31210989	1.20563071
14794	NM 014321	0.24655542	0.29586804	0.03729792	0.02796026	0.10705267	0.28907493	0.72249877	1.24038053	0.39751658	0.26385832	0.53326035	0.62054759	0.43487101
14795	NM 002720	0.03804582	0.20776376	0.01300807	0.05943485	0.08298822	0.2476458	0.11528011	0.1093043	0.27095562	0.25517699	0.50746173	0.36412117	0.13737491
14796	NM 007108	0.19574527	0.07856617	0.34602857	0.24013127	0.61298639	0.13807227	0.10303515	0.7051698	0.01773115	0.51954257	0.38942882	0.39377451	1.14207231
14797	NM 058219	0.54605854	0.59571558	0.33599138	0.24216484	0.58269382	0.60555208	0.46075016	0.24319802	0.37599277	0.19485338	0.26739922	0.50176054	0.39681831
14798	NM 015144	0.2720136	0.13058585	0.07252371	0.14146592	0.05882878	0.38051614	0.24498859	0.20094903	0.10104072	0.29002413	0.35623023	0.35646513	0.12641801
14799	NM 024946	0.02062072	0.14549266	0.1805163	0.03045222	0.11430232	0.02073941	0.05264947	0.20406014	0.233061	0.15463036	0.32041082	0.27608132	0.16769981
14800	NM 206824	0.31014276	0.07496439	0.26508132	0.2447052	0.0223187	0.02111789	0.50065392	0.06856018	0.26428938	0.12097725	0.05878961	0.16394927	0.09003771
14801	NM 002134	0.19297361	0.09994139	0.0057859	0.06368089	0.07680321	0.18500046	0.05817963	0.03882256	0.49282444	0.18758465	0.40981925	0.26384011	0.03524481
14802	AK027211	0.60768509	0.3043502	0.044475	0.14153405	0.07739827	0.20037904	1.10203052	1.13940477	0.27184957	0.12384019	0.21510521	0.22338234	0.1179161
14803	NM 001013839	0.21337169	0.23857792	0.18069701	0.10625174	0.06630493	0.08015756	0.26492366	0.49249029	0.33185402	0.0411216	0.0491272	0.5885632	0.0048041
14804	NM 002795	0.0937243	0.08370794	0.27204216	0.36185691	0.05005101	0.24851683	0.06814372	0.34713304	0.40216941	0.00251494	1.01013994	0.40293905	0.03177341
14805	NM 005993	0.25860235	0.31814161	0.29384747	0.23119402	0.04689999	0.03216252	0.23672442	0.14097282	0.65526402	0.6061008	0.2417087	0.20850998	0.91862141
14806	NM 001002033	0.01810829	0.09212466	0.21144836	0.12258144	0.67948776	0.0945081	0.10199711	0.69243896	0.78054696	0.24887091	0.19804776	0.00355606	0.5903971
14807	NM 020899	0.44362059	0.08774043	0.14920558	0.13584453	0.16220184	0.61288458	0.65594918	0.07697689	0.42609641	0.71906316	1.12807989	1.17673218	0.63357291
14808	NM 002615	0.33627349	0.20452915	0.25806865	0.26132932	0.37199327	0.30699125	0.57343298	0.48237777	1.04882777	0.42335466	0.12529223	0.39096242	0.47987581
14809	NM 014877	0.34046906	0.01221413	0.16107728	0.1004312	0.06034143	0.32608476	0.12231947	0.16652945	0.10069194	0.03947885	0.76019126	0.57715642	0.0607491
14810	NM 139033	0.25504383	0.00538092	0.00329315	0.01686166	0.0795799	0.08827583	0.13483676	0.40450063	0.37859264	0.44495878	0.45653459	0.62920511	0.28994191
14811	NM 000304	0.40089542	0.20824136	0.02166302	0.04475291	0.37560028	0.75648999	0.92012787	0.14998986	0.08833135	1.2320385	0.77952689	0.85872495	0.89729871
14812	NM 003726	0.16436491	0.37213937	0.07064062	0.33157071	0.19075324	0.77051342	0.79740411	0.25869182	0.35048455	1.32666437	0.07445441	0.79879821	0.30998831

Imagen 3.10. Valores absolutos ordenados según mercado (Muestras normales y tumorales)

Fuente: elaboración propia.

14764 genes 07 De Marzo 2013 sin Beta2 [Modo de compatibilidad] - Microsoft Excel

Archivo Inicio Insertar Diseño de página Fórmulas Datos Revisar Vista

Cortar Copiar Pegar Copiar formato Portapapeles Fuente Alineación Número Estilos

General Normal Buena Celda de co... Celda vincul... Ent

	N3	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI
14775															
14776															
14777															
14778															
14779		Unique id	CN2	CN8	CN54	CN42	CN83	CN85	CN78	MEDIA					
14780		NM 207446	0.52269149	0.05587619	0.18602891	0.13957353	0.46463871	0.49131879	0.01725937	0.268198142					
14781		NM 194272	0.33954856	1.02105045	0.27491599	0.87339729	0.1910193	0.21439877	0.87523091	0.541365896					
14782		NM 020371	0.0390427	0.10202263	0.16701977	0.20236959	0.03868085	0.10903247	0.2561394	0.130615342					
14783		NM 170589	0.14258635	0.59138739	0.14891638	0.26391885	0.1110931	0.31038031	0.1322366	0.242931282					
14784		NM 023076	0.25266778	0.43255889	0.03632422	0.09380069	0.07228161	0.06056086	0.13881926	0.155287617					
14785		NM 005153	0.02680916	0.00214372	0.174887	0.23738235	0.0005884	0.18020925	0.33980656	0.137403777					
14786		NM 001537	0.00263172	0.02287082	0.0150416	0.07655582	0.05240978	0.17622726	0.05763962	0.05762523					
14787		NM 024939	0.00705328	0.16106959	0.02187005	0.02263602	0.18632771	0.0308178	0.35179225	0.111652384					
14788		NM 004062	0.12972197	0.10016637	0.08193757	0.16109194	0.035396	0.35980186	0.22122291	0.155619802					
14789		NM 006159	0.41255161	0.44215357	0.33286658	0.70114595	0.09989033	0.31206182	0.26624885	0.366702673					
14790		NM 198390	0.31650195	0.14335504	0.20283952	0.31108066	0.08525592	0.12889239	0.0698864	0.179687413					
14791		AK024487	0.28438002	0.01200606	0.16780619	0.23574087	0.28322178	0.29098731	0.01297397	0.183873744					
14792		NM 020664	0.52226162	0.39818501	0.88059682	0.34470767	0.14905769	0.26069275	0.11964962	0.382164453					
14793		NM 0010394	0.25871184	0.3066082	0.08951968	0.27923825	0.19114614	0.10649594	0.15446772	0.198026824					
14794		NM 014321	0.24655542	0.29586804	0.03729792	0.02796026	0.10705267	0.28907493	0.72249877	0.24661543					
14795		NM 002720	0.03804582	0.20776376	0.01300807	0.05943485	0.08299822	0.2476458	0.11528011	0.10916809					
14796		NM 007108	0.19574527	0.07856617	0.34602857	0.24013127	0.61298639	0.13807227	0.10303515	0.24493787					
14797		NM 058219	0.54605854	0.59571558	0.33599138	0.24216484	0.58269382	0.60555208	0.46075016	0.481275199					
14798		NM 015144	0.2720136	0.13058585	0.07252371	0.14146592	0.05882878	0.38051614	0.24499859	0.185847513					
14799		NM 024946	0.02062072	0.14549266	0.1805163	0.03045222	0.11430232	0.02073941	0.05264947	0.080681873					
14800		NM 206824	0.31014276	0.07496439	0.26508132	0.2447052	0.0223187	0.02111789	0.50065392	0.205569169					
14801		NM 002134	0.19297361	0.09994139	0.0057859	0.06368089	0.07680321	0.18500046	0.05817963	0.097480729					
14802		AK027211	0.60768509	0.3043502	0.044475	0.14153405	0.07739827	0.20037904	1.10203052	0.35397888					
14803		NM 0010138	0.21337169	0.23857792	0.18069701	0.10625174	0.05630493	0.08015756	0.26492366	0.162897787					
14804		NM 002795	0.0937243	0.08370794	0.27204216	0.36185691	0.05005141	0.24851683	0.06814372	0.168291892					
14805		NM 005993	0.25860235	0.31814161	0.29384747	0.23119402	0.04689099	0.03216252	0.23672442	0.202509054					
14806		NM 0010020	0.01810829	0.09212466	0.21144836	0.12258144	0.67948776	0.0945081	0.10199711	0.188607961					
14807		NM 020899	0.44362059	0.08774043	0.14920558	0.13584453	0.16220184	0.61288458	0.65594918	0.321063818					
14808		NM 002615	0.33627349	0.20452915	0.25806865	0.26132932	0.37199327	0.30699125	0.57343298	0.330374017					
14809		NM 014877	0.34046906	0.01221413	0.16107728	0.1004312	0.06034143	0.32608476	0.12231947	0.160419618					
14810		NM 139033	0.25504383	0.00538092	0.00329315	0.01686166	0.0795799	0.08827583	0.13483676	0.083324579					
14811		NM 000304	0.40089542	0.20824136	0.02166302	0.04475291	0.37560028	0.75648999	0.92012787	0.389681549					
14812		NM 003726	0.16435491	0.37413937	0.07064062	0.33157071	0.19075324	0.77051327	0.79740411	0.385625196					

Imagen 3.11. Valores absolutos ordenados según mercado (Muestras de control)

Fuente: elaboración propia.



Imagen 3.12. Formación de la Cartera de genes.

Fuente: elaboración propia.

**2. Formación de carteras.** El valor de cada gen se divide entre el valor promedio de ese gen en las muestras de control restándole uno.

$$\text{Valor } r_{ij} = \frac{z_{ij}}{\bar{z}_{ic}} - 1$$

Donde:

$\bar{z}_{ic}$  es la media del gen i de los controles, de tal forma que  $r_{ij}$

representa una “rentabilidad”.



Copiar formato		Fuente		Alineación		Número		condicional como tabla		Estilos		
Portapapeles												
P2												
CARTELA DE GENES												
A	B	C	D	E	F	G	H	I	J	K	L	P
Uniqe id	CN2	CN8	CN54	CN42	CN83	CN85	CN78	CT 116	CT 89	CT 81	CT 3	
NO1 207446	0.94890048	0.71966079	0.30637509	0.47958081	0.73244567	0.81924627	0.93564694	0.69092826	0.37311173	0.61762892	2.51525228	2.59440878
NO1 194272	0.37279286	0.80606349	0.45218081	0.61332159	0.64715307	0.60396968	0.61670862	0.14701122	0.70660334	1.90368215	2.06062332	0.99020313
NO1 020371	0.70108644	0.2189076	0.27877419	0.54395654	0.70385678	0.16523991	0.96102076	0.7628449	0.73093512	0.37327825	2.88586857	0.45347951
NO1 170589	0.41305891	1.43438139	0.38700205	0.08639301	0.54269744	0.27764653	0.45566223	0.04230045	0.13332528	1.60085122	1.16234588	0.62070573
NO1 023076	0.62709551	1.78553732	0.76608424	0.39956514	0.53453073	0.61000847	0.10605066	5.11965661	0.83777696	0.0489356	0.18859009	0.31846901
NO1 005153	0.80488775	0.9843984	0.27279617	0.7272661	0.9957177	0.3153063	0.14730515	1.1566529	0.74411551	0.93373567	0.9771266	1.9078847
NO1 001537	0.95433025	0.603111	0.73897543	0.32851214	0.49005664	2.0581642	0.00249627	1.16273698	11.8421858	0.90842979	6.25691432	4.11308734
NO1 024939	0.36382821	0.44258973	0.40412378	0.77926345	0.66881982	0.73298442	2.15078132	0.43589018	0.63457513	0.97129719	0.88526613	1.16417375
NO1 004662	0.16641734	0.35633921	0.47347582	0.0351635	0.77254821	1.1205704	0.42156014	0.33699446	1.57130521	0.09766719	0.20908592	0.94852931
NO1 006159	0.12503028	0.20574598	0.09227119	0.91202846	0.72795858	0.14905088	0.27393807	1.37525233	0.76959574	0.612181	0.6963366	2.1199854
NO1 196390	0.76140299	0.20219764	0.12869459	0.7132236	0.52553204	0.28685447	0.6110668	0.2102572	0.8065342	0.07956499	0.10635687	0.70183659
AK024487	0.54660482	0.93470487	0.08736351	0.28208011	0.54300573	0.56253869	0.92944068	2.24027692	0.11823666	0.70826396	0.25413237	0.84104912
NO1 020664	0.36658869	0.0412926	1.30423528	0.09801222	0.60996454	0.31785193	0.68615859	0.74838264	0.50211338	0.26781976	0.9116999	0.32131443
NO1 01039476	0.30644849	0.5483165	0.54794165	0.41010318	0.3474625	0.46221485	0.21996569	0.55173193	1.35807137	1.66003154	1.52645348	0.57609904
NO1 014321	0.60024334	0.19971422	0.8487608	0.88662403	0.56591253	0.17216862	1.24956762	0.42961443	0.61188851	0.06991812	1.16235141	1.51625165
NO1 002720	0.65149322	0.90315469	0.88084362	0.45556699	0.23972091	1.26848154	0.05598721	0.1024773	1.48203868	1.33746898	3.64844378	2.33541762
NO1 007108	0.20083740	0.67924041	0.41271975	0.01962374	1.50261989	0.36429678	0.57934168	1.87897415	0.72670961	1.21114984	0.59890859	0.6075058
NO1 058219	1.3460768	0.23778575	0.30187264	0.45626689	0.2107895	0.25822415	0.04264719	0.49467992	0.21875227	0.59513106	0.44439435	0.04256472
NO1 051144	0.94353866	0.29745499	0.60976768	0.23880651	0.63845671	1.07446247	0.31827747	0.80125756	0.45634266	0.56048989	0.91677817	0.91805166
NO1 029456	0.74441937	0.80328806	0.23738364	0.62256427	0.4701388	0.74294833	0.34744361	1.52919435	1.88861443	0.91654404	0.7120651	0.42185077
NO1 206824	0.90070268	0.63532353	0.2894938	0.1903789	0.89142971	0.08972711	1.43454239	0.664681	0.28564697	0.41150099	0.71405141	0.20246179
NO1 002134	0.97960782	0.0252451	0.94064566	0.3467335	0.21211907	0.89975176	0.40316786	0.70147117	0.55200803	0.92435251	3.20410525	1.70658379
AK027111	0.71672697	0.14020238	0.87435691	0.60016246	0.78134777	0.43382613	1.21132669	2.21884959	0.23010754	0.65014808	0.39231992	0.36893897
NO1 01013839	0.30985017	0.65458661	0.1092662	0.34779983	0.65435423	0.50792729	0.62631836	0.20230867	0.3719168	0.74561594	0.69841705	2.61308277
NO1 007295	0.4430849	0.56026826	0.61648997	1.15017433	0.7025917	0.47670113	0.59086051	1.0626848	1.3897135	0.98505606	0.0230901	1.39426677
NO1 005993	0.27699155	0.57099943	0.45103375	0.14164782	0.76844991	0.46117996	0.16895122	0.30386904	0.23272703	0.19956454	0.19356982	0.02963289
NO1 00102033	0.0393988	0.51155475	0.12109985	0.3500728	0.26642625	0.49891774	0.49921012	0.27131363	1.23854274	0.9351434	0.10504983	0.98114577
NO1 02089	0.38172092	0.72671966	0.5327575	0.57689244	0.49479777	0.90891825	1.0430492	0.76024428	0.3271393	1.23862629	2.51356895	2.65510369
NO1 002615	0.01785696	0.38016335	0.21885512	0.20898948	0.12597816	0.07077666	0.73570848	0.46009596	2.17466874	0.28144053	0.62075562	0.18339337
NO1 048877	1.1223655	0.9238614	0.0049961	0.73349692	0.62385256	1.03628979	0.23750302	0.8008654	0.73232151	0.75390261	0.73978743	0.9779204
NO1 139033	2.0608475	0.93542217	0.9604781	0.79763877	0.9449409	0.0541222	0.61821116	0.85451758	0.3453589	0.4306063	0.47899902	0.55125457
NO1 000304	0.028777	0.46561414	0.9440841	0.88515616	0.03613512	0.94130308	0.61320207	0.16059933	0.77332427	2.16165469	1.00042032	1.20565823
NO1 003726	0.5739625	0.02974895	0.81681533	0.14017362	0.5054302	0.99808888	1.0678216	0.32916227	0.09112645	2.44037323	0.80689954	0.13738262
NO1 160616	0.9388144	1.0484083	0.92330478	0.0151907	0.83355967	0.38330676	1.1186721	0.10510056	0.8416293	0.59512173	0.30526191	0.27022498
NO1 01040431	0.12707776	0.16332105	0.21777935	0.05026261	0.07170638	0.6341268	0.74647397	0.59159486	6.23086527	0.35996494	0.39195538	3.49622801
NO1 01033580	0.10594479	0.58464205	0.51950801	0.18624487	0.17538469	2.35993273	0.95526769	0.28068073	1.64146869	0.27675032	3.79420048	0.30225544
NO1 051355	0.17374544	0.37689327	0.2108699	0.77194532	0.36160475	0.34089828	0.05094523	0.04833077	0.48201811	0.74973835	0.45253723	1.16961357
NO1 182538	0.40753769	0.6672115	1.06972609	0.70830123	0.96771763	0.70971111	0.96346353	0.84823392	1.6199105	0.10211415	1.74798393	0.86728451
NO1 178493	0.73526414	0.59881622	0.6778693	0.04323435	0.24219383	0.06061682	0.52200345	3.41535324	2.5351271	0.7352548	0.87945964	0.45160514
NO1 007267	0.41427904	0.14237222	0.67864393	0.97688785	0.5816077	0.25178271	0.8212832	0.45615044	0.92347971	0.16960222	0.49193367	3.11593106
NO1 01040078	0.3988511	0.9579276	0.77891428	0.70131869	0.897451	1.0987393	1.1889227	2.44433063	1.43766725	1.56895333	0.63398543	1.14683727
NO1 016429	0.28582126	0.41238105	0.82309869	0.56949087	0.85963593	0.85736557	1.26830012	0.14328185	0.3007727	1.53473111	0.66638702	0.75887107
NO1 004715	0.0103606	0.85563769	0.39665457	0.16913262	0.50439365	1.53877287	0.41680897	1.37476194	0.13601681	0.9647354	0.17252698	0.17507501
NO1 002091	0.62702186	0.90032328	0.8047319	0.5790944	0.64793202	0.84092226	0.31044688	1.71319759	0.4282117	0.67314204	0.9414334	0.04563072
NO1 004539	0.6436365	0.25256494	0.25984707	0.91120478	0.28156793	0.8727105	0.99792542	8.30684639	0.95849606	3.34890958	1.16814142	0.56081318
NO1 01008467	1.2681455	0.40256984	0.29711551	0.6963384	0.1422888	0.83290523	1.67041938	9.91078813	1.31682840	0.72561239	0.36341755	0.85253706
												0.46077974
												0.44458718
												0.1263791
Datos Beta Plántila Resumen Filtro con Beta 1 Datos originales con filtro bet												

- $\bar{d}_c$  representa la media para las muestras de control.

En este caso  $R_j$  representa la variabilidad, en tanto por uno, de la muestra j respecto a las muestras de control.

**3. Obtención de las betas.** El siguiente paso es calcular los parámetros estadísticos necesarios para el posterior análisis: media, mediana, varianza, desviación típica y la covarianza entre cada uno de los genes y la fila Mercado.

Este estudio pretende realizar un ajuste lineal entre la variable  $r_i$  que son las variabilidades definidas anteriormente en el gen i en las 96 muestras y R que son las variabilidades totales de cada una de las 96 muestras.

Utilizando el modelo de regresión lineal simple de mínimos cuadrados ordinarios (M.C.O.), la  $r_i$  es la variable respuesta (variable dependiente) y R la variable explicativa (variable independiente). Dicha función queda expresada de la siguiente manera:

$$r_i = \beta_i R + \alpha_1$$

Donde:

- R: variable que recoge el sumatorio de las variabilidades de cada muestra
- $\beta_i$ : beta del gen i

- $r_i$ : gen i

El valor de la beta significa:

- Si  $\beta_i > 1$ : la variabilidad de ese gen es mayor que la variabilidad del tumor.
- Si  $\beta_i = 1$ : que la variabilidad del gen aumenta o disminuye en la misma medida que lo hace la variabilidad de las muestras.
- Si  $\beta_i = 0$ : que no existe relación entre la variabilidad del gen y la variabilidad de las muestras.

Interesa identificar aquellos genes que tengan una variabilidad significativa superior a la de la muestra pues, ello indicaría, que la expresión de esos genes en las muestras tumorales es más alta o más baja (dependiendo de si está sobre-expresado o inhibido) que en las muestras del pool.

En la Hoja “Beta Plantilla” se realizan los cálculos de las betas de los 14764 genes. Para su cálculo se han tomado tanto las muestras tumorales como las muestras de control. Se filtran los genes en función de su beta, reteniendo aquellos genes cuya beta es superior a un determinado valor.

A continuación, en una nueva tabla llamada “DATOS REDUCIDOS” se seleccionan los genes cuya beta era mayor que 3.6. En la tabla “DATOS

COMPARADOS” se recoge el valor de los genes aportados por el hospital en valor absoluto. Se realiza el sumatorio de los genes incluidos esta vez en el análisis de cada una de las muestras.

The image shows a screenshot of a Microsoft Excel spreadsheet. The title bar indicates the file is '14764 genes 28 De JULIO 2014 DEFINITIVO [Modo de compatibilidad] - Microsoft Excel'. The spreadsheet has a grid with columns labeled CY, CZ, DA, DB, DC, DD, DE, DF, DG, DH, DI, DJ, DK, DL, DM, DN, DO. The rows contain numerical data, with some cells highlighted in blue and others in yellow. The title 'DATOS REDUCIDOS' is visible in the center of the table. The spreadsheet also shows the standard Excel ribbon with tabs like Archivo, Inicio, Insertar, etc.

Imagen 3.14. Imagen de la tabla “DATOS REDUCIDOS”.

Fuente: elaboración propia.



14764 genes 28 De JULIO 2014 DEFINITIVO [Modo de compatibilidad] - Microsoft Excel

ArchivoInicioInsertarDiseño de páginaFórmulasDatosRevisarVista

Cortar

Copiar

Copiar formato

Portapapeles

Arial

10

Ajustar texto

Combinar y centrar

General

%000%

Formato condicional

Dar formato como tabla

Normal

Buena

Incorrecto

Neutral

Cálculo

Celda de co...

Celda vincul...

Entrada

Notas

Salida

Fuente

Alineación

Numero

Estilos

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
14775																	
14776																	
14777																	
14778																	
14779	Unique id	CN2	CN8	CN54	CN42	CN83	CN85	CN78	CT 116	CT 89	CT 81	CT 3	CT 98	CT 29	CT 82	CT 115	CT 12
14780	NM_207446	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14781	NM_194272	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14782	NM_020371	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14783	NM_170585	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14784	NM_023076	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14785	NM_005153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14786	NM_001537	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14787	NM_024939	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14788	NM_004062	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14789	NM_005159	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14790	NM_198390	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14791	AK024487	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14792	NM_020664	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14793	NM_001039476	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14794	NM_014321	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14795	NM_007120	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14796	NM_007108	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14797	NM_058219	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14798	NM_015144	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14799	NM_024946	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14800	NM_206824	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14801	NM_007134	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14802	AK027211	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14803	NM_01013839	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14804	NM_002795	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14805	NM_005993	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14806	NM_001002033	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14807	NM_020899	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14808	NM_002515	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14809	NM_014877	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14810	NM_139033	0.25504383	0.00538092	0.00329315	0.01686166	0.0795799	0.08827583	0.13483676	0.40450063	0.37859264	0.44495878	0.45653459	0.62920511	0.28994191	0.31216061	0.21016839	0.3437362
14811	NM_000304	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14812	NM_003726	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14813	NM_016016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14814	NM_001040431	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14815	NM_001033580	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14816	NM_015355	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14817	NM_182538	0.05353318	0.03007488	0.18704604	0.02636151	0.17782721	0.15446183	0.00330189	0.01046205	0.2367675	0.18093579	0.2483418	0.16875091	0.3651948	0.10184041	0.3264218	0.13725197
14818	NM_178493	0.1108468	0.02562742	0.1071816	0.06111141	0.04840784	0.06344347	0.0305339	0.28204799	0.22582009	0.01691164	0.24781568	0.3283706	0.9646436	0.07772359	0.14851569	0.23830691
14819	NM_007267	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14820	NM_001040078	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14821	NM_016429	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14822	NM_014425	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Imagen 3.15. Imagen de la tabla “DATOS COMPARADOS”.

Fuente: elaboración propia.

Se calcula valor promedio de cada uno los genes en las muestras de control. Calculand0 los  $r'_{ij}$ .

$$r'_{ij} = \frac{z'_{ij}}{\bar{z}'_{ic}} - 1,$$

donde:

- $z'_{ij}$  es la media del gen i, cuya beta es superior a 3.6, de los controles, de tal forma que  $r'_{ij}$  representa una “rentabilidad”.
- $\bar{z}'_{ic}$  es la media de los controles de los genes cuya beta es superior a 3.6

Del mismo modo se halla  $R'_j$ :

$$R'_j = \frac{d'_j}{\bar{d}'} - 1 \quad \text{de éstos genes,}$$

Donde:

- $d'$  es el sumatorio de genes cuya beta ha es mayor a 3.6 de una muestra
- $\bar{d}'$  representa la media para las muestras control de los genes cuya beta es superior a 3.6.

En este caso  $R'_j$  representa la variabilidad, en tanto por uno, de la muestra j respecto a las muestras control.

14764 genes 07 De Marzo 2013 sin Beta2 [Modo de compatibilidad] - Microsoft Ex

ArchivoInicioInsertarDiseño de páginaFórmulasDatosRevisarVista

Cortar

Copiar

Copiar formato

Imagen 3.15. Formación de la Cartera de genes.

Fuente: elaboración propia.

### 3.3.2. Análisis

Una vez calculados los datos se procede al análisis de los mismos. Los datos obtenidos son:

La tabla “CARTERA DE GENES” donde se recogen:

- $r_{ij}$ : las “rentabilidades” de los 14.764 genes objeto de estudio.



- $R_j$  las “rentabilidad de Mercado” que corresponde, como ya se ha explicado anteriormente, al sumatorio de las rentabilidades de los genes en cada una de las 96 muestras.

Parámetros estadísticos de todos los genes:

- Medidas de tendencia central: media y mediana.
- Parámetros de dispersión: varianza y desviación típica.
- La covarianza y la correlación que existe de cada uno de los genes con “el Mercado”.

Las 14.764 rectas de regresión de cada uno de los genes con “el Mercado”, para calcular las 14.764 betas de los genes.

14764 genes 28 De JULIO 2014 DEFINITIVO [Modo de compatibilidad] - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
29548																	
29549																	
29550	Uniqued Id	CN2	CN8	CN54	CN42	CN83	CN85	CN78	CT 116	CT 89	CT 81	CT 3	CT 98	CT 29	CT 82	CT 115	CT 12
29551	NM 267446																
29552	NM 194272																
29553	NM 020371																
29554	NM 170589																
29555	NM 023076																
29556	NM 005153																
29557	NM 001537																
29558	NM 024939																
29559	NM 004062																
29560	NM 006159																
29561	NM 198390																
29562	AK024487																
29563	NM 020664																
29564	NM 001039476																
29565	NM 014321																
29566	NM 002720																
29567	NM 007108																
29568	NM 058219																
29569	NM 015144																
29570	NM 024946																
29571	NM 206824																
29572	NM 002134																
29573	AK027211																
29574	NM 001013839																
29575	NM 002795																
29576	NM 005993																
29577	NM 001002033																
29578	NM 020899																
29579	NM 002615																
29580	NM 014877																
29581	NM 139033	2.06084756	-0.93542217	-0.9604781	-0.79763877	-0.0449409	0.05942122	0.61821116	3.85451758	3.543589	4.3400663	4.47899072	6.55125457	2.47966843	2.7463209	1.52228567	3.12526779
29582	NM 000304																
29583	NM 003726																
29584	NM 016016																
29585	NM 001040431																
29586	NM 001033580																
29587	NM 015355																
29588	NM 182538	-0.40763769	-0.6672115	1.06972609	-0.70830123	0.96771673	0.70917111	-0.96346353	-0.88423392	1.6199105	1.00211415	1.74798393	0.86728451	3.04100091	0.12689778	2.61196487	0.51873836
29589	NM 178493	0.73526414	-0.59881262	0.67788693	-0.04332435	-0.24219383	-0.00681682	-0.52200345	3.41535324	2.53512711	-0.7352548	2.87945964	4.14051603	14.1011262	0.21673297	1.32495621	2.73060343
29590	NM 007267																
29591	NM 001040078																
29592	NM 016429																
29593	NM 004715																
29594	NM 002091																
29595	NM 004539																

Imagen 3.16. Imagen de la tabla “CARTERA REDUCIDA”

Fuente: elaboración propia.

La tabla “CARTERA REDUCIDA” recoge:

$r'_{ij}$  las “rentabilidades” de los 787 genes (son los genes del estudio cuya beta es superior a 3.6).

$R'_j$  la “rentabilidad de Mercado” que corresponde, como ya se ha explicado anteriormente, al sumatorio de las rentabilidades de los 787 genes en cada una de las 96 muestras.

Parámetros estadísticos de los 521 genes:

- Medidas de tendencia central: media y mediana.
- Parámetros de dispersión: varianza y desviación típica
- La covarianza y la correlación que existe de cada uno de los genes con “el Mercado”.

Las 521 rectas de regresión de cada uno de los genes con “el Mercado” para calcular las 521 betas de los genes.

El objetivo que se persigue es identificar aquellos genes que contribuyen más a la variabilidad del tumor. Para tratar de identificar esos genes se comparan los datos de las columnas “SUMATORIO” y “SUMATORIO REDUCIDO”. En la tabla “DATOS VALORES ABSOLUTOS ORDENADOS” de la hoja “DATOS” donde se recogen los valores de todos los genes (14764). En la última fila de la tabla, llamada “SUMATORIO” se ha calculado la suma de cada una de las muestras. Esta fila representa la variabilidad que tienen las muestras sin ningún tipo de filtrado de genes. En la tabla “DATOS COMPARADOS” se recogen los datos de los genes cuya beta supera un determinado valor en cifras absolutas. En la última fila de la tabla, llamada “SUMATORIO REDUCIDO” se ha calculado la suma de cada una de las muestras. Esta fila representa la variabilidad que tienen las muestras del nuevo conjunto de genes.

14764 genes 07 De Marzo 2013 sin Beta2 [Modo de compatibilidad] - Microsoft Ex

ArchivoInicioInsertarDiseño de páginaFórmulasDatosRevisarVista

Cortar

Copiar

Copiar formato

Pegar

Arial

10

A<sup>+</sup>A<sup>-</sup>

Ajustar texto

Combinar y centrar

General

%

000

Formato condicional

Dar formato como tabla

Normal

Buena

Inc

Celda de co...

Celda vincul...

Ent

PortapapelesFuenteAlineaciónNúmeroEstilos

B3

</

Imagen 3.17.

Fuente: elaboración propia.

Al realizar el cociente, muestra a muestra, entre el valor de la variabilidad de la muestra reducida<sup>60</sup> y el valor de la variabilidad de la muestra total<sup>61</sup> se obtiene el porcentaje de variabilidad para cada una de las muestras que aportan los genes seleccionados (los que tienen una beta superior al valor fijado). A este cociente le llamamos VF<sub>j</sub>. Interesa saber cuál es la variabilidad

<sup>60</sup> Dichos valores están recogidos en la hoja "BETA" en la última fila de la tabla "DATOS REDUCIDOS"

<sup>61</sup> Dichos valores están recogidos en la hoja "DATOS" en la última fila de la tabla "DATOS VALORES ABSOLUTOS ORDENADOS"

final media de las muestras controles ( $VF_C$ ), cuál es la variabilidad final media de las muestras tumorales ( $VF_T$ ) y cuál es la variabilidad final media, de manera conjunta, tanto de las muestras tumorales como de las muestras de control ( $VF_{C+T}$ ).

$VF_C = 0,01642$  es decir un 1,642%.

$VF_T = 0,05794$  es decir un 5,794%.

CUADRO RESUMEN			
BETA1 1,4 MEDIANA + 1,7 DES TIP			
Genes			14.764
GENES TOTAL	14764	GENES SI (análisis)	521
PORCENTAJE	0,03528854	GENES NO (suprimo)	14254
0,46535225		1,632093	
GENES	521	GENES	14764
Med(CNredu/CN14764)	0,0164216	Med(CTredu/CT14764)	0,057594

Imagen 3.18. Imagen de la tabla “RESUMEN DE LOS DATOS”

Fuente: elaboración propia.

En la tabla “DATOS REDUCIDOS” de la hoja “Beta plantilla” únicamente a parecen los datos de los genes cuya beta es superior a la beta establecida. Esta tabla es necesaria para continuar con la selección de genes.

Imagen 3.19. Imagen de la tabla “DATOS REDUCIDOS”

Fuente: elaboración propia.

En función de si los genes están presentes o no en esta tabla “DATOS REDUCIDOS” de la hoja “Beta Plantilla” seleccionaremos los datos originales de las muestras. Es decir, si un gen tiene una beta superior a la beta marcada, en este trabajo la beta considerada es 1.4, los datos de los genes utilizados para la elaboración de la tabla “CARTERA DE GENES” aparecen en esta tabla.

En la hoja “Filtrado con Beta 1” creamos una nueva tabla llamada “SELECCIÓN DE GENES EN FUNCIÓN DE LA BETA 1”. Si los datos de un gen figuraban en la tabla anterior en esta nueva tabla escogeremos los datos originales, sin ninguna transformación, proporcionados por el hospital, tal y como se refleja en la siguiente imagen.



14754 genes 30 De JULIO 2014 92 Muestras con Filtro de media [Modo de compatibilidad] - Microsoft Excel

ArchivosInicioInsertarDiseño de páginaFórmulasDatosRevisarVista

<

Imagen 3.20. Imagen de la tabla “SELECCIÓN DE GENES EN FUNCIÓN BETA”

Fuente: elaboración propia.

A continuación se calculan algunos datos como:

- Media
- Mediana
- Desviación típica
- Curtosis

Se establece un intervalo para cada uno de los genes cuyo límite inferior será la Mediana menos una coma cinco veces la desviación típica y su límite superior la Mediana más una coma cinco veces la desviación típica.

$$\text{Intervalo } (Me - 1.5\sigma, Me + 1.5\sigma)$$

Establecemos ahora una nueva condición para todos los genes en cada una de las muestras: si el dato original del gen, está fuera de este intervalo, entonces aparece dicho dato en la última tabla “SELECCIÓN DE GENES EN FUNCIÓN MEDIANA”

Archivo

Inicio

Diseño de página

Fórmulas

Referencias

Vista

Inicio

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Referencias

Imagen 3.21. Imagen de la tabla “SELECCIÓN DE GENES EN FUNCIÓN MEDIANA”

Fuente: elaboración propia.

Las celdas que contienen datos se verán, para una mejor referencia visual, en rojo, tal y como puede verse en la siguiente imagen 3.22.



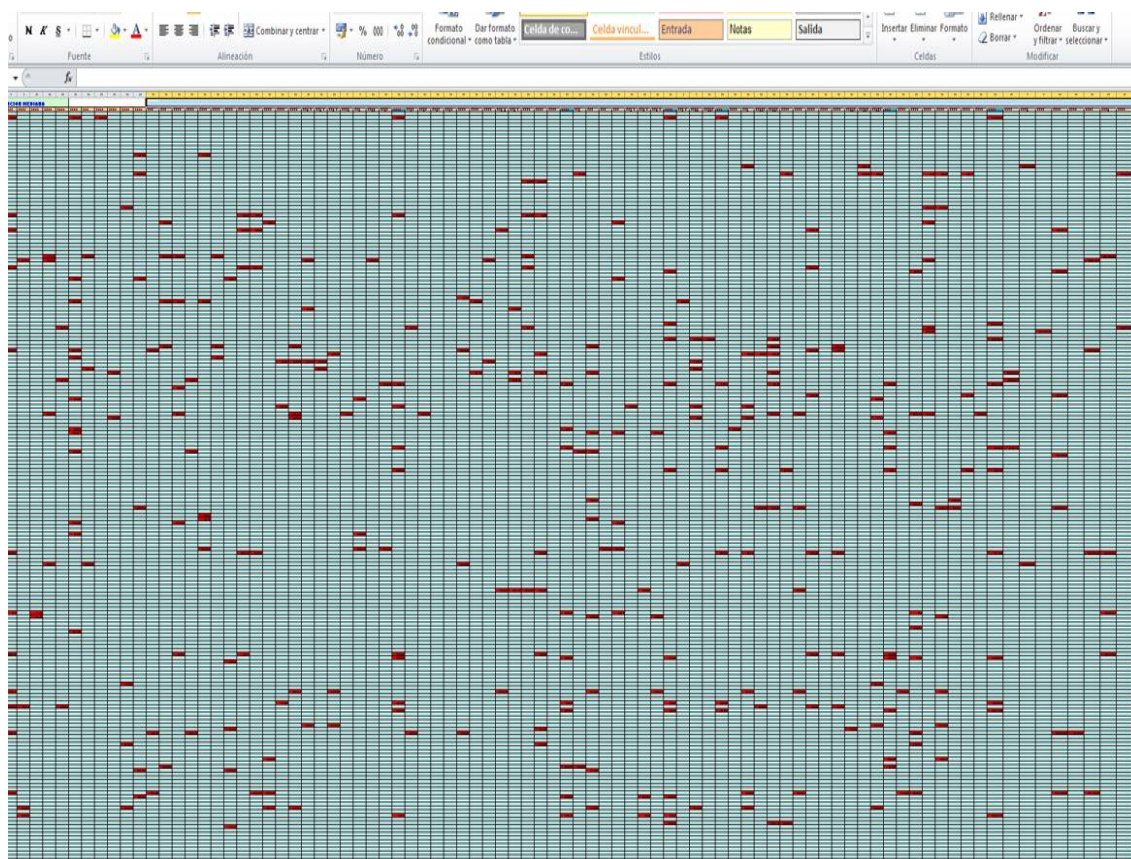


Imagen 3.22. Imagen de la tabla “SELECCIÓN DE GENES EN FUNCIÓN MEDIANA”

Fuente: elaboración propia.

Para finalizar este filtrado se exige que el gen esté fuera del intervalo seleccionado  $(Me - 1.5\sigma, Me + 1.5\sigma)$  en, al menos, 8 muestras. Con esto me aseguro que los genes seleccionados no sólo aparezcan en las muestras de control (6 muestras) sino como mínimo, en dos muestras tumorales también.

El siguiente paso es elaborar una clasificación de los genes seleccionados. Para ello utilizamos el software Genesping 12.6.

Escojo los genes datos cuya beta sea superior al valor fijado y el gen este presente fuera del intervalo, tal y como se ha explicado, en al menos 8

muestras. Estos datos quedan recogidos en la hoja “Filtrado con Beta 1” en dos columnas: la columna del identificador ID y la columna que hace referencia a la superación, o no, superación, de las dos condiciones expuestas. Dicho columna puede tomar dos valores:

Uno: ese gen forma parte de la selección final de genes realizada al haber complicado ambas condiciones.

Cero: ese gen no forma parte de la selección final de genes realizada al no haber cumplido una o las dos condiciones establecidas.

GU	GV	GW	GX	GY	GZ	HA	HB	HC	HD	HE	HF
condición	contar				8						
						Beta	1,4				
						mediana	1,65		Ge. Coinciden		
contar	596	Unique id	suma		14764	genes si		1722		0	
0	0	NM_207446	0		NM_207446	#N/A	1	NM_194272	#N/A		
8	0	NM_194272	0		NM_194272	#N/A	2	NM_004062	#N/A		
0	0	NM_020371	0		NM_020371	#N/A	3	NM_002615	#N/A		
0	0	NM_170589	0		NM_170589	#N/A	4	NM_003726	#N/A		
0	0	NM_023076	0		NM_023076	#N/A	5	NM_178493	#N/A		
0	0	NM_005153	0		NM_005153	#N/A	6	NM_002091	#N/A		
0	0	NM_001537	0		NM_001537	#N/A	7	NM_003012	#N/A		
0	0	NM_024939	0		NM_024939	#N/A	8	NM_000064	#N/A		
0	0	NM_004062	0		NM_004062	#N/A	9	NM_144947	#N/A		
0	0	NM_006159	0		NM_006159	#N/A	10	NM_152358	#N/A		
0	0	NM_198390	0		NM_198390	#N/A	11	NM_002842	#N/A		
2	0	AK024487	0		AK024487	#N/A	12	NM_001828	#N/A		
0	0	NM_020664	0		NM_020664	#N/A	13	NM_003500	#N/A		
0	0	NM_0010394	0		NM_0010394	#N/A	14	NM_006887	#N/A		
3	0	NM_014321	0		NM_014321	#N/A	15	NM_006770	#N/A		
0	0	NM_002720	0		NM_002720	#N/A	16	NM_0010085	#N/A		
10	1	NM_007108	1		NM_007108	#N/A	17	AF435956	#N/A		
0	0	NM_058219	0		NM_058219	#N/A	18	NM_014585	#N/A		
2	0	NM_015144	0		NM_015144	#N/A	19	NM_003881	#N/A		

Imagen 3.23. Imagen de donde aparecen los genes seleccionados en función de las dos condiciones establecidas.

Fuente: elaboración propia.

En la columna condición figuran dos valores: uno y cero. El valor uno indica que la beta de dicho gen es superior a la beta marcada como condición para efectuar la selección. El valor cero indica que la beta de ese gen tiene un valor menor o igual a la beta fijada para realizar la selección de genes.

En la tabla “DATOS REDUCIDOS” DE LA HOJA “Beta plantilla” únicamente aparecen los datos de los genes cuya beta es superior a la beta establecida.

	A	B	C	D
7	NM_005153	0		
8	NM_001537	0		
9	NM_024938	0		
10	NM_004063	0		
11	NM_006159	0		
12	NM_198390	0		
13	AK024487	0		
14	NM_020664	0		
15	NM_001039	0		
16	NM_014321	0		
17	NM_002721	0		
18	NM_007108	0		
19	NM_058219	0		
20	NM_015144	0		
21	NM_024341	0		
22	NM_206824	0		
23	NM_002134	0		
24	AK027211	0		
25	NM_001013	0		
26	NM_002739	0		
27	NM_005393	0		
28	NM_001002	0		
29	NM_020893	0		
30	NM_002619	0		
31	NM_014877	0		
32	NM_139033	2,060848		
33	NM_000304	0		
34	NM_003721	0		
35	NM_016016	0		
36	NM_001040	0		
37	NM_001033	0		
38	NM_015359	0		
39	NM_182538	0		
40	NM_178493	0,735264		
41	NM_007261	0		
42	NM_001040	0		
43	NM_016429	0		
44	NM_004719	0		
45	NM_002093	0		
46	NM_004533	0		
47	NM_001080	0		
48	NM_001914	0		
49	NM_003012	-0,92536		
50	NM_031446	0		
51	AF086436	0		
52	NM_025041	0		
53	NM_017908	0		
54	NM_032821	0		
55	NM_012155	0		
56	ENST00000	0		
57	NM_023003	0		
58	NM_032881	0		
59	NM_001001	0		
60	NM_024106	0		
61	NM_030621	0		
62	NM_198540	0		
63	NM_024074	0		
64	ENST00000	0		
65	NM_000064	0		

Imagen 3.24. Imagen del nuevo libro de Excel

Fuente: elaboración propia.

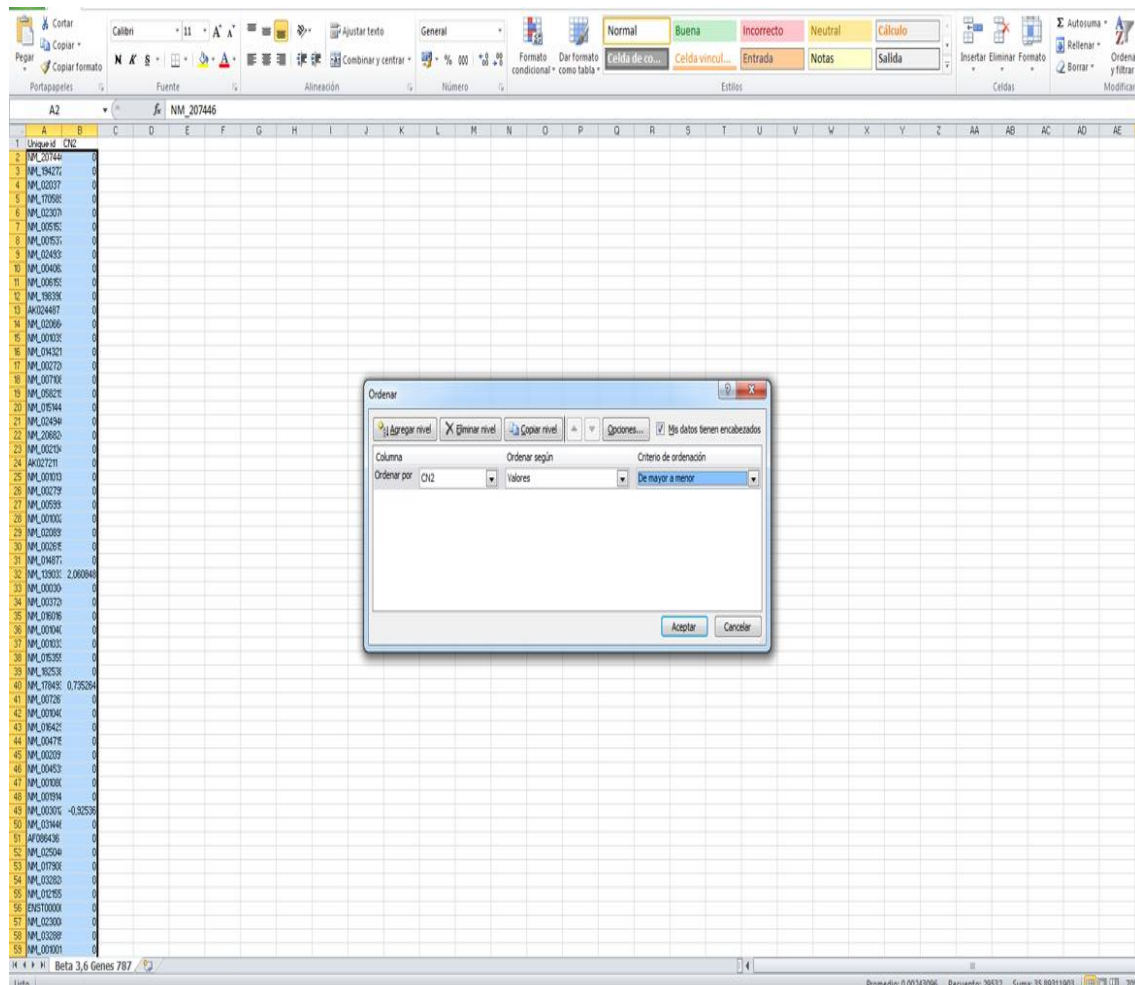


Imagen 3.25. Imagen del proceso para ordenar los datos.

Fuente: elaboración propia.

Elimino todos los genes cuyo valor, en esta muestra, sea igual a 0 (lo que supone prescindir, tal y como se ha explicado, de los datos que no satisfacen la condición).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
334	BC039021	0,0455248															
335	THC2563227	0,0446224															
336	NM_004052	0,0438491															
337	ENST000002	0,0424668															
338	BX640843	0,0380724															
339	NR_003088	0,036721															
340	NM_003567	0,0318473															
341	NM_079423	0,0243943															
342	THC2756581	0,0192467															
343	NM_016548	0,017968															
344	BE540788	0,0129844															
345	ENST000003	0,0115314															
346	NM_005663	0,0113991															
347	NM_001009	0,0084572															
348	THC2532456	0,0051919															
349	NM_001063	0,001596															
350	NM_153498	0,0007006															
351	NM_000014	0,000317															
352	NM_207446	0															
353	NM_194272	0															
354	NM_020371	0															
355	NM_170589	0															
356	NM_023076	0															
357	NM_005153	0															
358	NM_001537	0															
359	NM_024939	0															
360	NM_004062	0															
361	NM_006159	0															
362	NM_198390	0															
363	AK024487	0															
364	NM_020664	0															
365	NM_001039	0															
366	NM_014321	0															

Imagen 3.26. Imagen del proceso de eliminación de los datos.

Fuente: elaboración propia.

Finalmente, elimino también la columna de los valores pues ya tengo la selección de genes realizada.

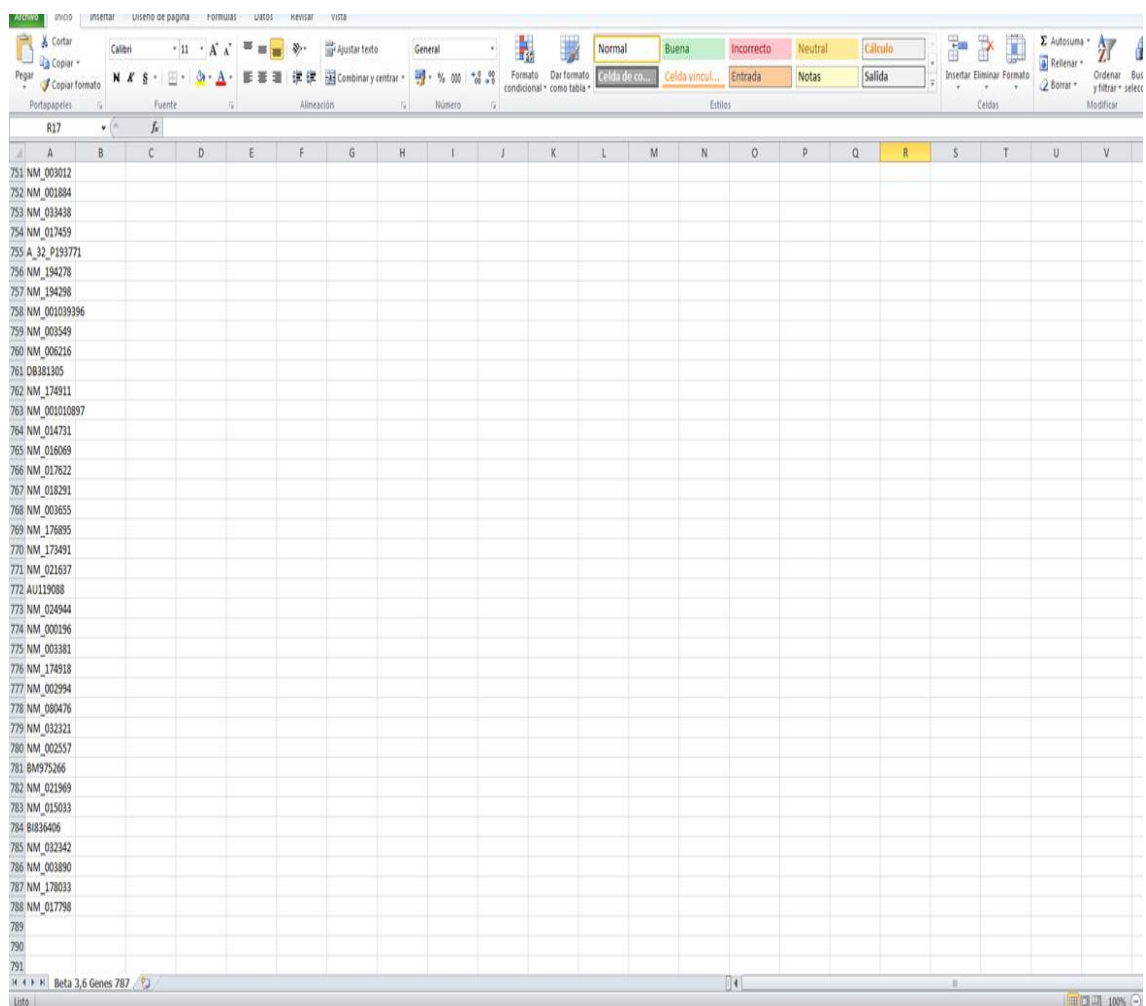


Imagen 3.27. Imagen de los genes seleccionados.

Fuente: elaboración propia.

Para que los datos puedan ser leídos por la aplicación GeneSpring es necesario guardar el archivo de Excel como texto (delimitado por tabulaciones).

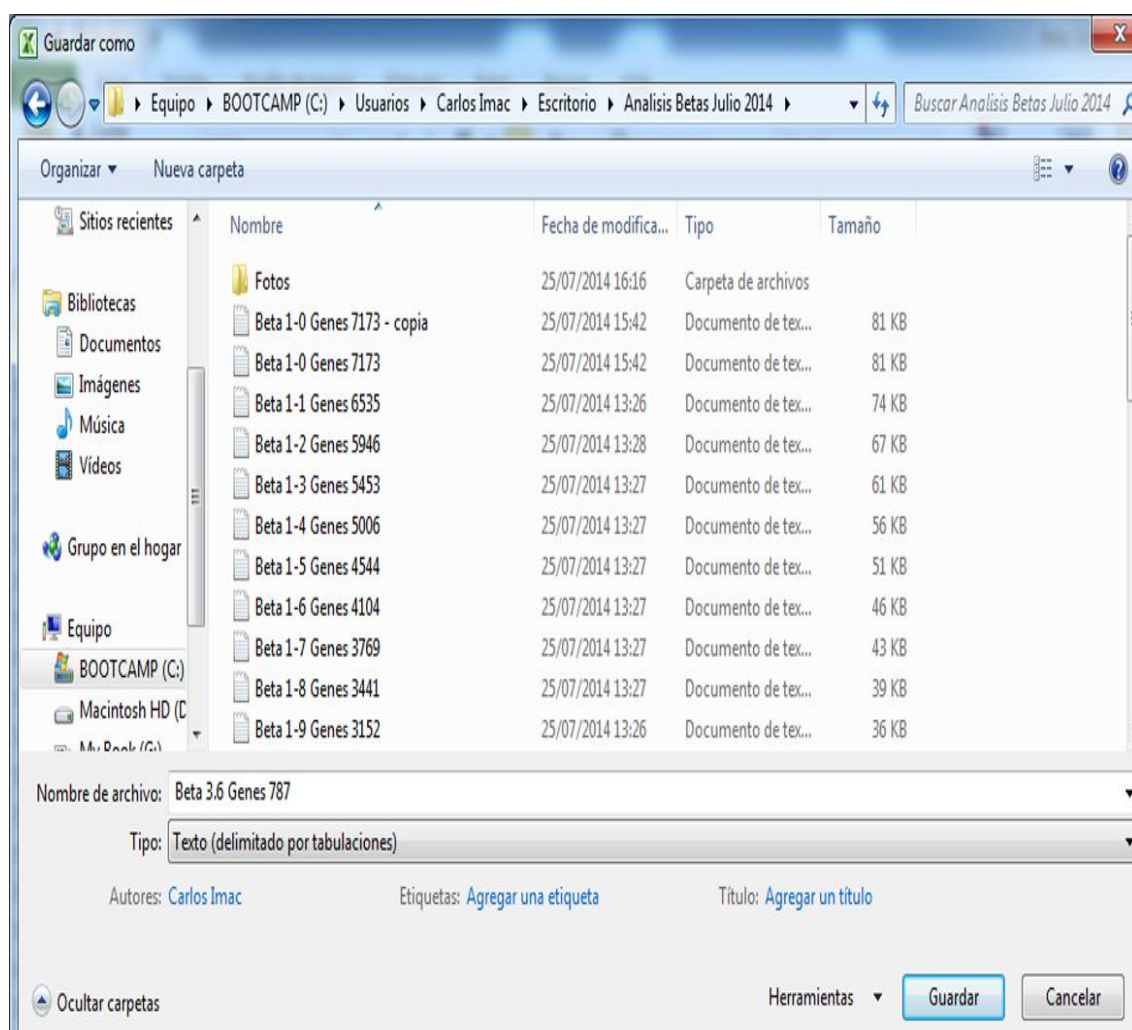


Imagen 3.28. Imagen de los datos guardados.

Fuente: elaboración propia.

Abrimos el programa GeneSpring 12.6 y cargamos los datos facilitados por el Laboratorio de Oncología Molecular del Hospital Clínico San Carlos de Madrid pertenecientes a los 14764 genes de las 96 muestras.



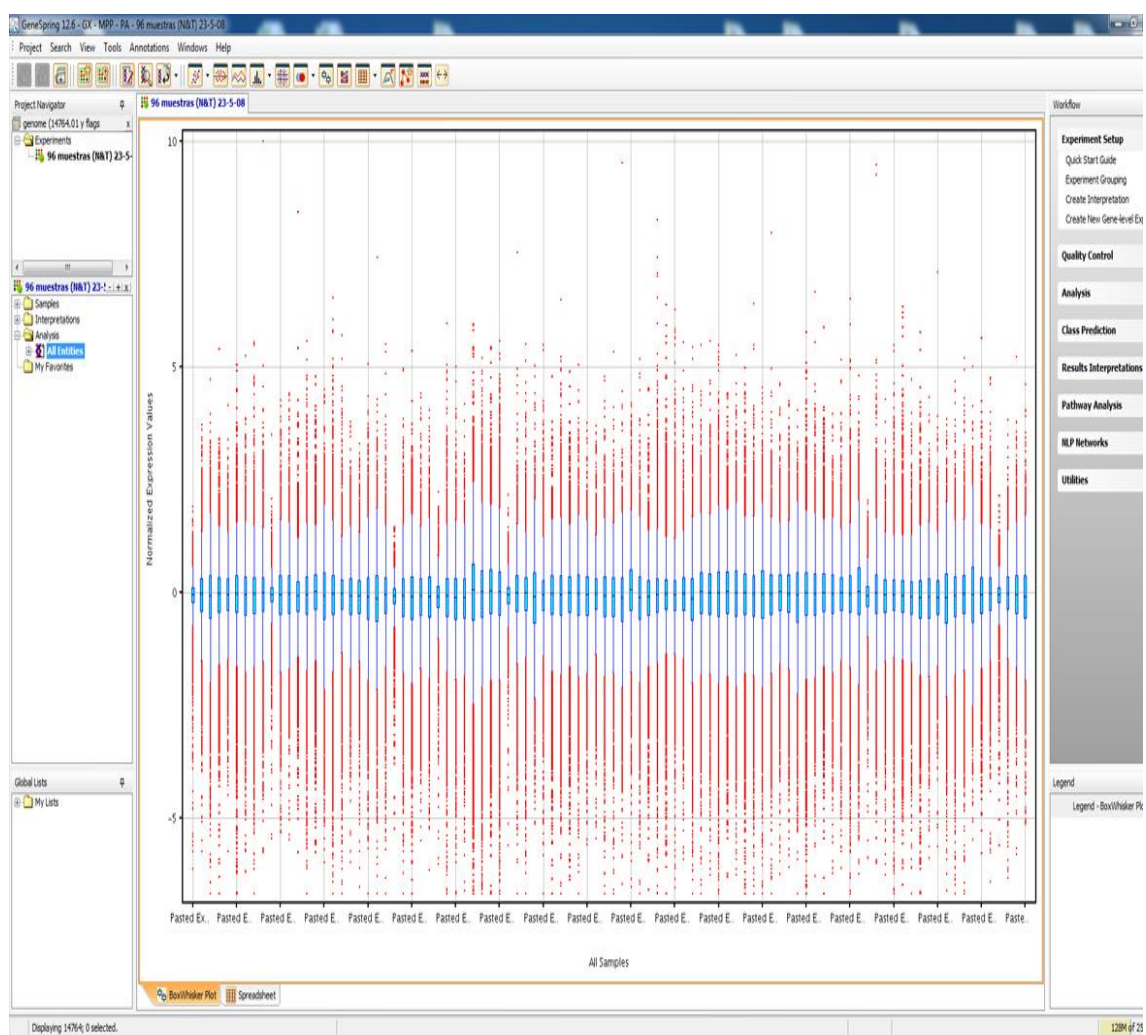


Imagen 3.29. Imagen de la aplicación Genespring 12.6.

Fuente: elaboración propia.

A continuación importo el archivo que contiene los genes cuya beta ha superado la condición establecida.

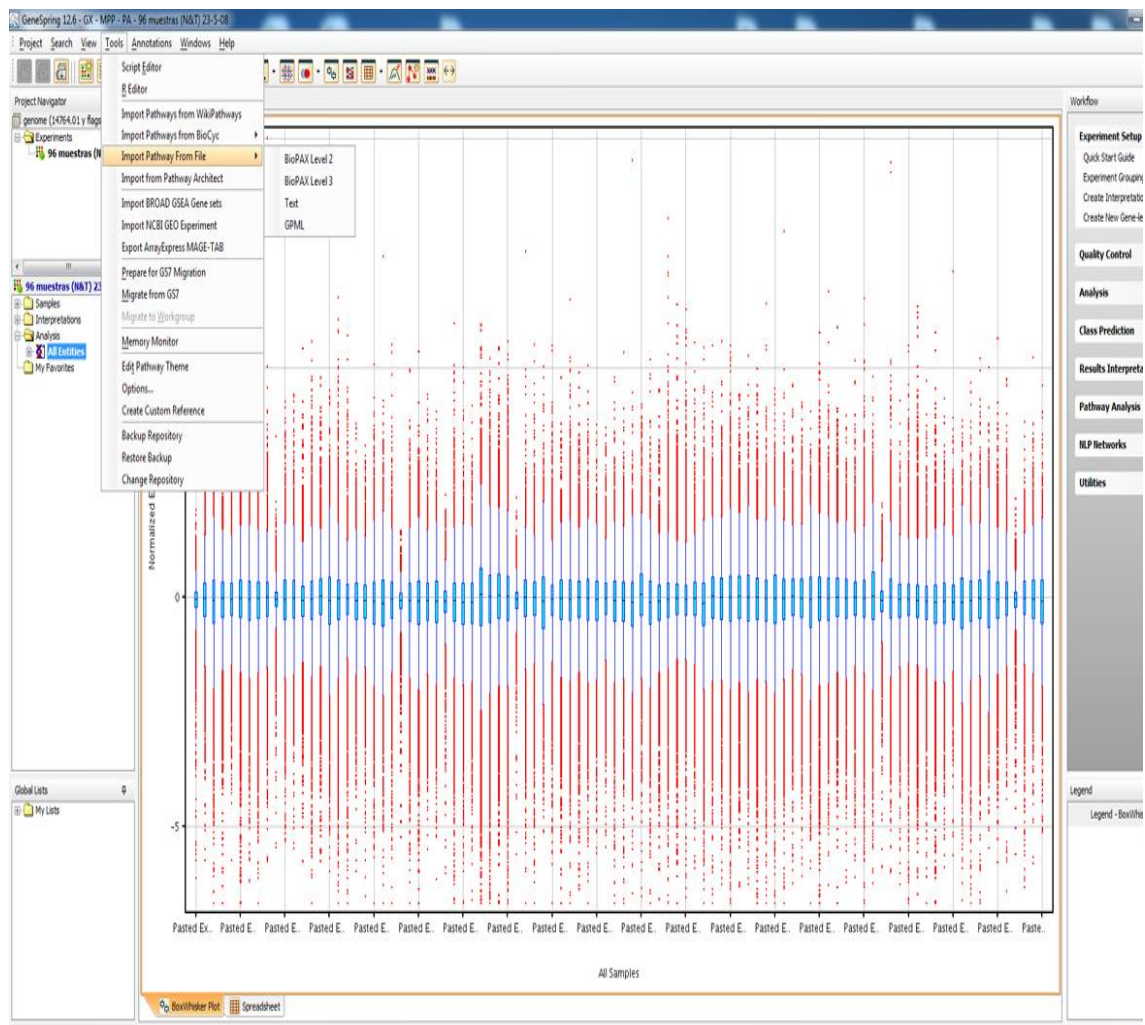


Imagen 3.30. Imagen del proceso para cargar los datos del hospital en la aplicación Genespring 12.6.

Fuente: elaboración propia.

En la parte de la izquierda de la interface del programa selecciono, dentro de “All Entities” la carpeta “Imported List” e importo la lista con los 787 genes cuya beta es superior a 3.6

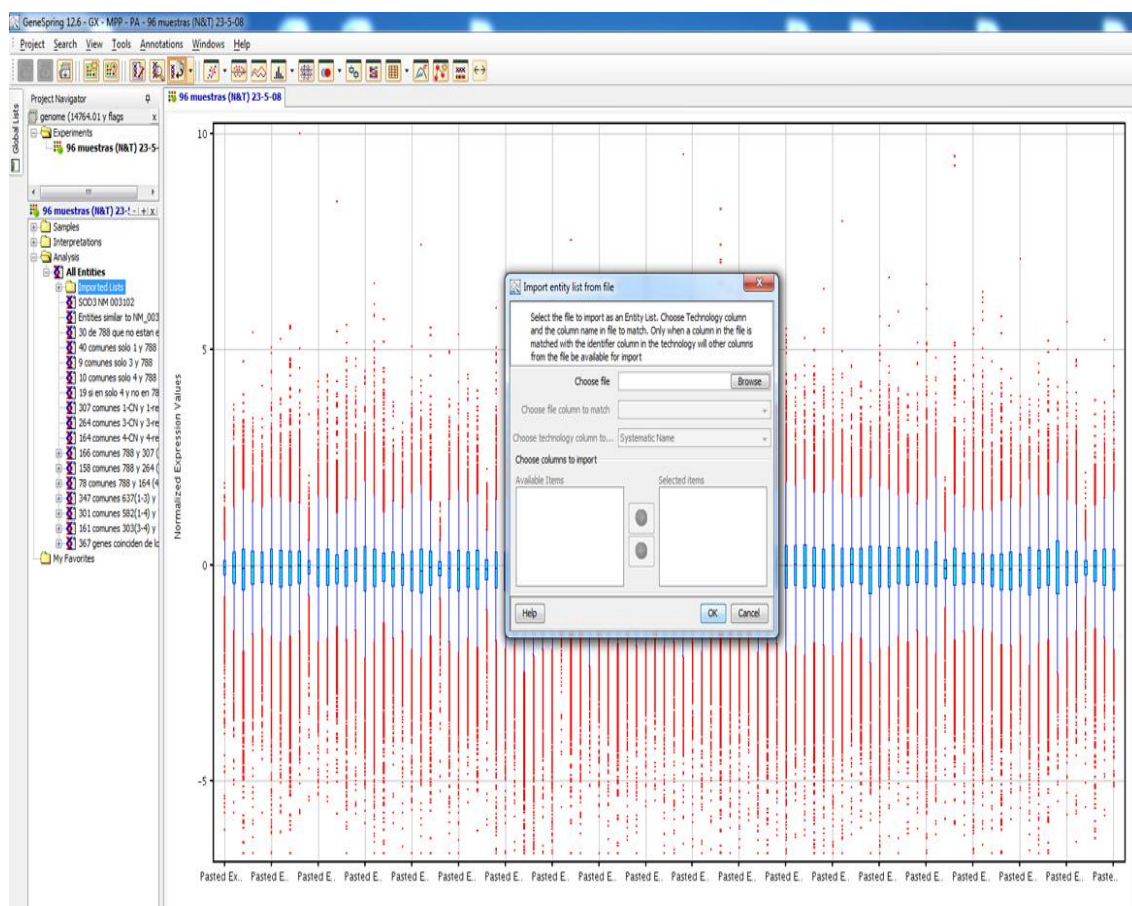


Imagen 3.31. Imagen del proceso para importar los datos en la aplicación Genespring 12.6.

Fuente: elaboración propia.

Realizo un clúster jerárquico utilizando la distancia euclídea y como método de agrupación jerárquica el método de Ward<sup>62</sup>, similar al que emplean en el laboratorio. Para ello en la parte derecha de la interface del programa selecciono la opción Analysis – Clustering con la siguiente configuración:

<sup>62</sup> Es un procedimiento jerárquico en el cual, en cada etapa, se unen los dos clústers para los cuales se tenga el menor incremento en el valor total de la suma de los cuadrados de las diferencias, dentro de cada clúster, de cada individuo al centroide del clúster.

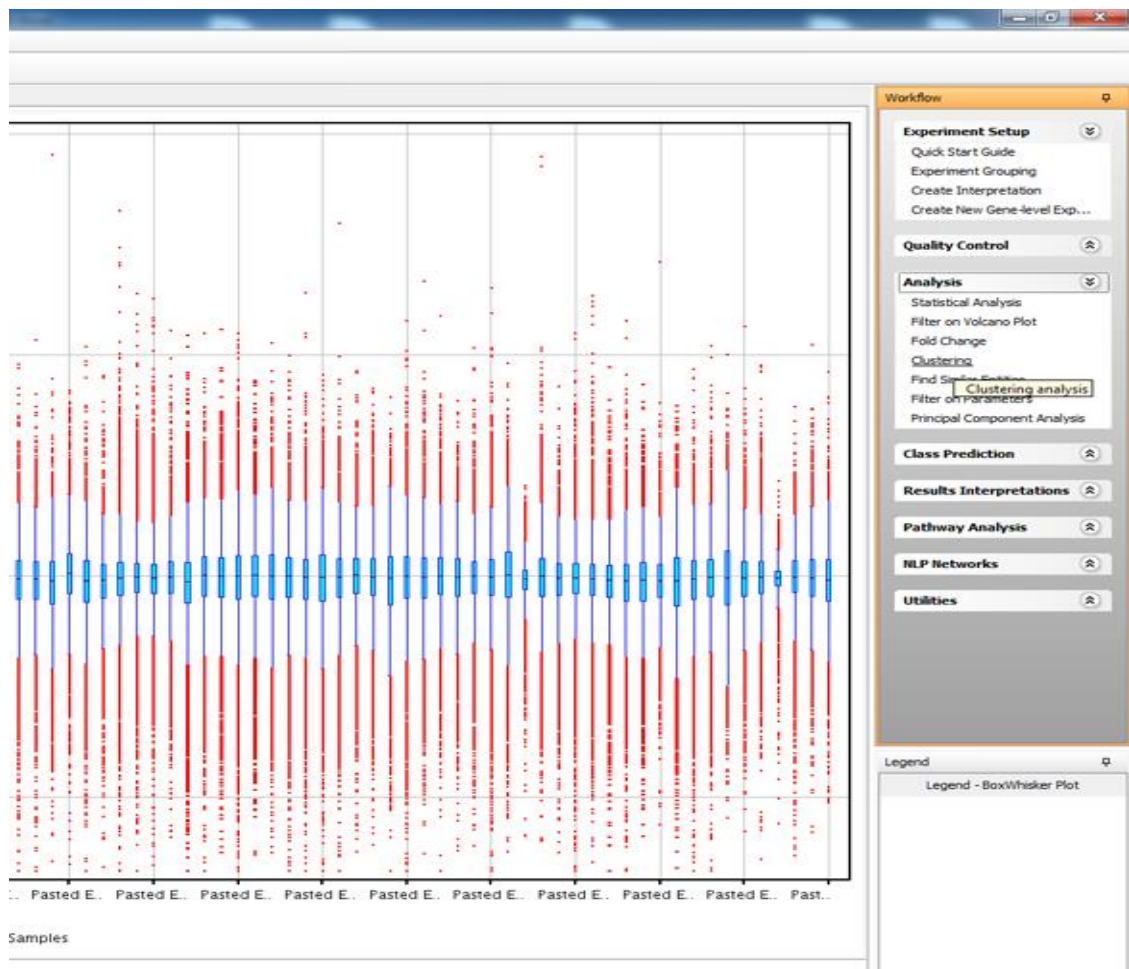


Imagen 3.32. Imagen del proceso para elaborar un clúster en la aplicación Genespring 12.6.

Fuente: elaboración propia.

Se hace clic con el botón izquierdo del ratón en *Entity list* para cargar la lista de genes filtrados según la condición de la beta.

Interpretation: Se seleccionan los datos que utiliza el laboratorio para la elaboración del clúster.

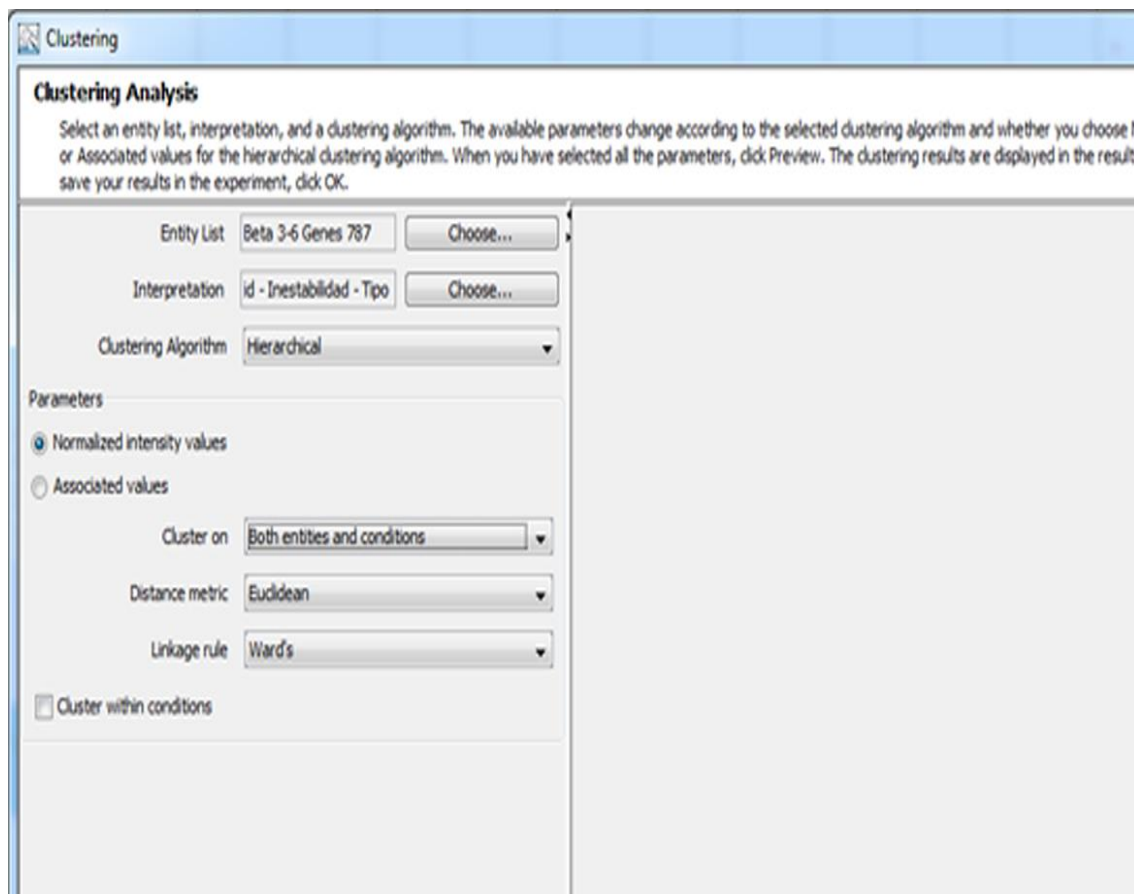


Imagen 3.33. Imagen del proceso para elaborar un clúster en la aplicación Genespring 12.6.

Fuente: elaboración propia.





### 3.3.3. Resultados

Los 521 genes cuyas betas son superiores a 1,4 y, al menos, en 9 muestras están fuera del intervalo (Mediana - 1.7 Desviaciones típicas, Mediana + 1.7 Desviaciones típicas) representan el 3,5280% del conjunto de datos, tal y como puede verse en la siguiente tabla.

CUADRO RESUMEN			
BETA1 1,4 MEDIANA + 1,7 DES TIP			
Genes			14.764
GENES TOTAL	14764	GENES SI (análisis)	521
PORCENTAJE	0,03528854	GENES NO (suprimo)	14254
0,46535225		1,632093	
GENES	521	GENES	14764
Med(CNredu/CN14764)	0,0164216	Med(CTredu/CT14764)	0,057594

Imagen 3.35.

Fuente: elaboración propia.

Esto significa que el conjunto de formado por los 521 genes explican el 5,759% de la variabilidad en las muestras tumorales. Si todos los genes contribuyeran de manera alícuota a la variabilidad total del tumor, el porcentaje que le correspondería a cada gen sería del 0,006320% y al grupo de los 521 genes sería del 3,5280%. Como puede apreciarse, el valor de esta selección de genes es 1,66672 veces más.

En cambio, estos mismos genes explican el 1,6422% de la variabilidad de las muestras de control 0,46535 veces de lo que explicarían un grupo de 521 genes si la variabilidad se repartiera, como hemos dicho anteriormente, de manera alícuota entre todos los genes de las muestras. Es decir, estos 521 genes explican 3,5288 veces más la variabilidad media en las muestras tumorales que en las muestras de control.

Si los datos iniciales que nos proporciona el Hospital los ordenamos en función del sumatorio de cada muestra<sup>63</sup> obtenemos una distribución de las muestras en cuanto a su variabilidad, considerando los 14.764 genes. Como puede observarse en la imagen 3.35 las muestras de los controles quedan todas a la izquierda. Esto verifica la hipótesis inicial: cuánto menos variabilidad tengan las muestras mejor.

---

<sup>63</sup> Se calcula el valor absoluto de cada uno de los genes para evitar que las variabilidades positivas se anulen con las variabilidades negativas. Finalmente se suma el valor absoluto de cada una de las muestras.



The image shows a screenshot of a Microsoft Excel spreadsheet. The title bar at the top indicates the file name is "VALORES ABSOLUTOS ORDENADOS". The spreadsheet is organized into columns labeled with gene identifiers (e.g., CT16, CT17, CT18, etc.) and rows containing numerical data. The data is sorted in ascending order of absolute values. The interface includes standard Excel toolbars for editing, formatting, and calculation. The formula bar at the top shows the active cell's content, which appears to be a numerical value. The spreadsheet is displayed in a window titled "Excel - VALORES ABSOLUTOS ORDENADOS".

Imagen 3.36. En la tabla “VALORES ABSOLUTOS ORDENADOS” puede observarse como quedan ordenadas las muestras teniendo en cuenta la variabilidad de todos los genes. Los controles, cuyo identificador aparece en azul, son los que menos variabilidad presentan y están situados a la izquierda de la tabla.

Fuente: elaboración propia.

Al realizar el mismo proceso una vez se tienen los 521 se obtiene una distribución muy similar a la inicial; lo que indica que los genes seleccionados son representativos del total de genes ya que la muestra escogida distribuye al conjunto de datos como lo hace toda la población (los 14764 genes).





De los 521 genes obtenidos en este estudio, 306 genes coinciden con los genes seleccionados por el grupo de la Dra. Perez-Villamil, lo que representa, aproximadamente, el 60% de los genes.

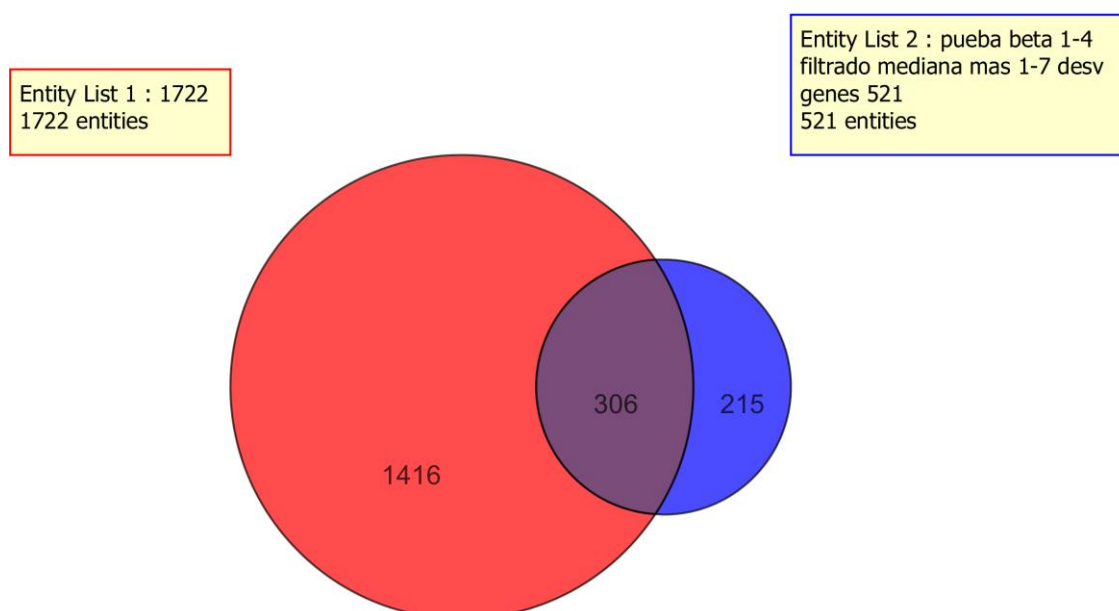


Imagen 3.38. Gráfico que muestra el número de genes que comparten ambos estudios. El gráfico rojo corresponde con los genes del grupo de la Dra. Pérez-Villamil y el gráfico azul son los genes obtenidos en el estudio de esta Tesis<sup>65</sup>.

Fuente: elaboración propia.

Al realizar la asociación de genes mediante relaciones jerárquicas para agrupar los subtipos de cáncer de colon en grupos, en función de la expresión de los 521 genes, se han obtenido 6 cluster para los siguientes subtipos de cáncer de colon:

<sup>65</sup> En el anexo 3 se recoge con detalle tanto los genes que comparten ambos estudios como los genes específicos seleccionados en esta Tesis.

- Grupo 1: cáncer colorectal que presenta bajo nivel de estroma.
- Grupo 2: cáncer colorectal relacionado con inmunoglobinas (no se ha podido aún corroborar)
- Grupo 3: cáncer colorectal que presenta alto en de estroma.
- Grupo 4: cáncer colorectal inestable mucionoso.

Se ha obtenido la siguiente agrupación, mostrada en las imágenes 3.39 y 3.40.

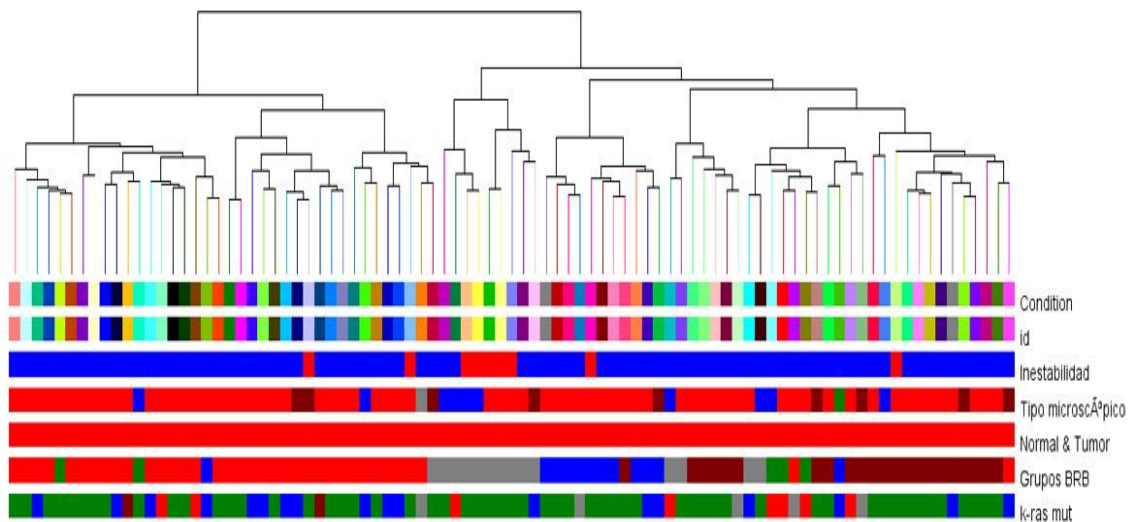


Imagen 3.39. Gráfico del cluster de los 521 genes.

Fuente: elaboración propia.

En la Imagen 3.40 observamos los 6 cluster con diferentes colores.

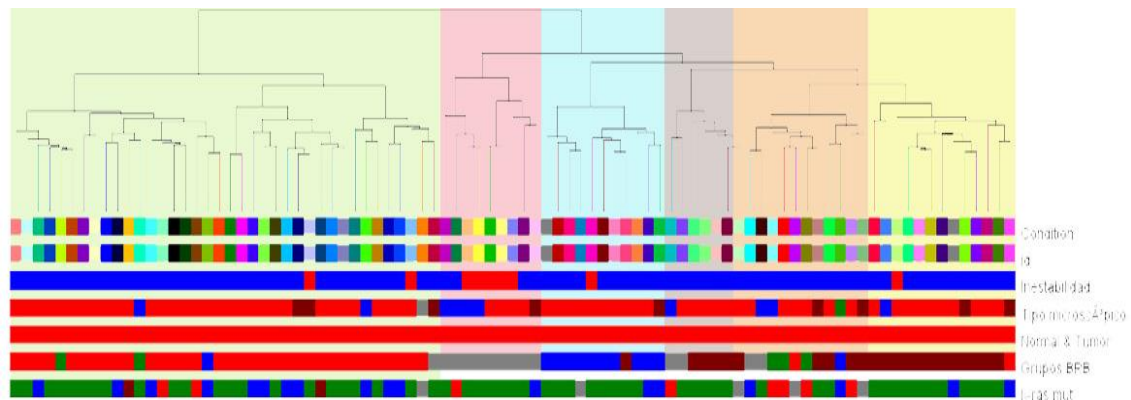


Imagen 3.40. Gráfico del cluster

Fuente: elaboración propia.

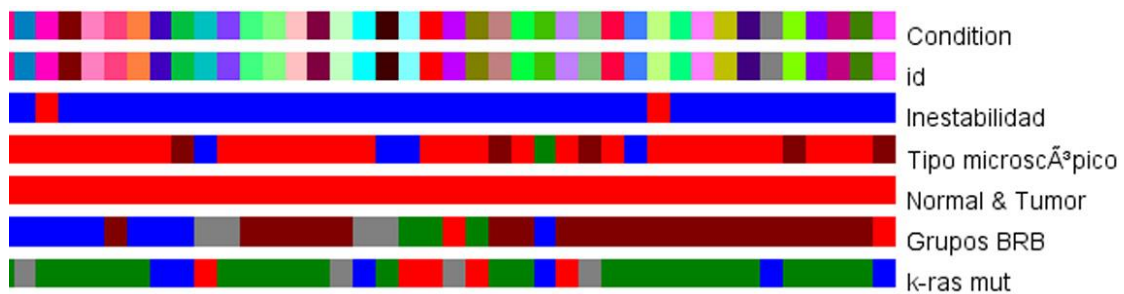


Imagen 3.41. Sección de la imagen del cluster.

Fuente: elaboración propia.

Las filas: *Condition* e *ID* de la anterior imagen nos indican la referencia con las que se han nombrado a las muestras. La fila *Normal & Tumor* indica si la muestra es normal o tumoral (Color azul tumoral, color rojo normal). La fila *Grupos BRB* recoge la información referente al subtipo de cáncer (el grupo 1 es el color rojo, grupo 2 color azul, grupo 3 color marrón y grupo 4 color gris). La fila *k-ras mut* revela si el gen *k-ras* presenta mutación y el tipo de ésta.

Como puede observarse en la Imagen 3.42, el clúster de color verde califica por completo al subgrupo 1 de cáncer colorrectal, el que presenta bajos niveles de estroma. El clúster de color azul clasifica al subtipo 2, cáncer colorrectal relacionado con las inmunoglobinas. El clúster de color rojo clasifica al subtipo 3, cáncer colorectal que presenta alto en de estroma.Finalmente, el clúster de color amarillo clasifica al subtipo 4, cáncer de colon inestable mucinoso

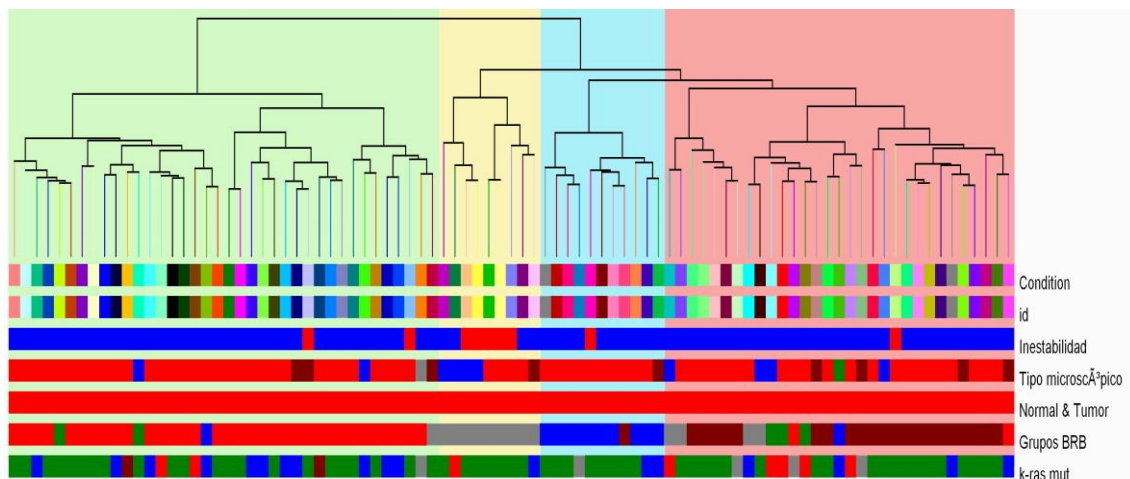


Imagen 3.42: Gráfico del clustering efectuado.

Fuente: elaboración propia.

Al incluir las muestras normales en el cluster obtenemos lo siguiente:

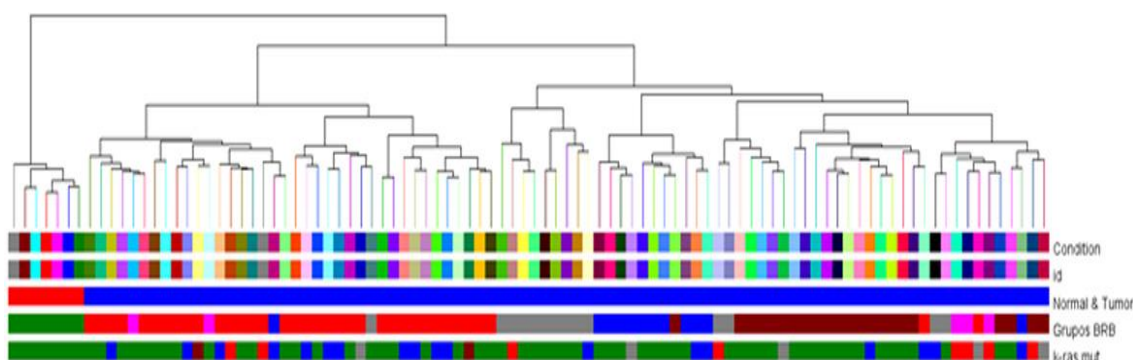


Imagen 3.42. Imagen del cluster incluyendo las muestras normales.

Fuente: elaboración propia.

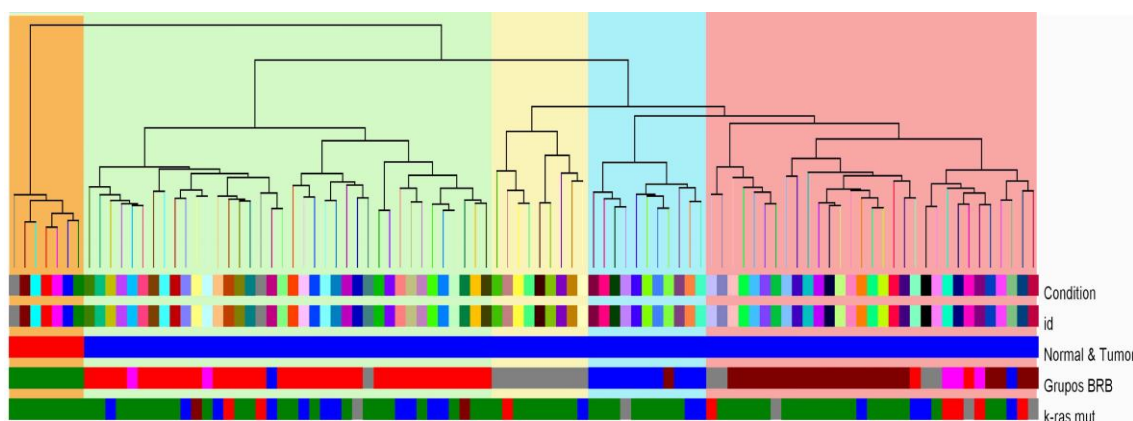


Imagen 3.43. Imagen del cluster incluyendo las muestras normales.

Fuente: elaboración propia.

Como puede apreciarse en las filas *Tumor & Normal* de las imágenes 3.42 y 3.43 todas las muestras normales, como era de esperar, quedan agrupadas bajo un mismo cluster (color amarillo oscuro en la Imagen 3.43.).

## **CAPÍTULO IV. CONCLUSIONES, IMPLICACIONES Y RECOMENDACIONES**

### **4.1. CONCLUSIONES**

El progreso tecnológico ha contribuido en los últimos años a profundizar en los estudios moleculares del cáncer permitiendo, mediante el filtrado de genes y su posterior clustering, la identificación de biomarcadores predictivos que se traducen en estrategias personalizadas para numerosos subtipos de cánceres. El objetivo del presente estudio es emplear el parámetro beta, utilizado en la medición del riesgo en el análisis bursátil, en la selección de genes que presenten patrones de expresión diferencialmente expresados entre las muestras sanas y las tumorales en el cáncer colorrectal y aportar, en el caso de que sea válido, una técnica más para la identificación de biomarcadores predictivos.

La selección de genes se ha realizado en base a su variabilidad (midiendo cuánto varía la expresión de un gen respecto de un supuesto estado normal). Sumando todas las variabilidades de los genes presentes en cada una de las muestras obtenemos la variabilidad de todas las muestras. Con ello tratamos de buscar patrones de expresión que clasifiquen subgrupos tumorales.

De los 14764 genes, 521 tienen una beta mayor a 1.4 y están fuera del intervalo (Mediana – una como siete desviaciones típicas, Mediana + una como siete desviaciones típicas), en almenos, más de 8 muestras. Este conjunto de



genes representa un 3.5288% respecto al total de genes. Si la variabilidad de una muestra se repartiera alícuotamente entre todos los genes de esa muestra; es decir, si todos los genes de una muestra tienen aportan la misma variabilidad a dicha muestra, el valor que aportara este conjunto de genes a la variabilidad total de las muestras debería ser del 3.52% y, como se ha constatado, esto no ocurre. La variabilidad que este conjunto de 521 genes seleccionados aporta a las muestras tumorales y normales es de 5,7594%, y 1,1642% respectivamente. La variabilidad que aportan estos genes a las muestras tumorales es 1,6320 veces más de lo que le correspondería si la variabilidad se repartiera de forma alícuota entre todos los genes. A la vista de estos datos, parece clara la importancia de estos 521 genes.

Al analizar más detenidamente los datos se comprueba que la variabilidad que este grupo de genes aporta a las muestras normales es del 1,6421% frente al 3,5288% que cabría esperar. Es decir, explican menos de la mitad de variabilidad de lo que, si la contribución de cada gen fuera la misma, en las muestras normales, concretamente 0.4653 veces menos.

Al comparar la variabilidad que aportan los 521 genes en las muestras tumorales frente a las muestras normales se observa que es 3,5072 veces, más del triple.

Se ha tomado como hipótesis de partida que la no variabilidad tiene una relación directa con la malignidad del tumor. De tal manera que, cuanto menos variabilidad presente una muestra, menos malignidad cabe esperar que tenga dicha muestra. Esta hipótesis se verifica tanto al ordenar las muestras en función de la variabilidad de sus genes como en la posterior agrupación

jerárquica de las muestras. Las muestras que menos variabilidad presentan son, precisamente, las muestras normales, tal y como se refleja en las imágenes 4.1 y 4.2.

O

Copiar formato										Combinar y centrar										Formato condicional como tabla										Celda vinculada										Entera										Notas										Borrar										Ver y filtrar																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
Portapapeles										Fuente										Alineación										Número										Estilos										Celdas										Modific																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
Q38																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			

Tabla 4.1.Clasificación de las muestras según su variabilidad<sup>66</sup>.

Fuente: elaboración propia

<sup>66</sup> Para ver más información sobre la tabla consultar en el documento de Excel la tabla “VALORES ABSOLUTOS ORDENADOS” de la hoja “Datos”

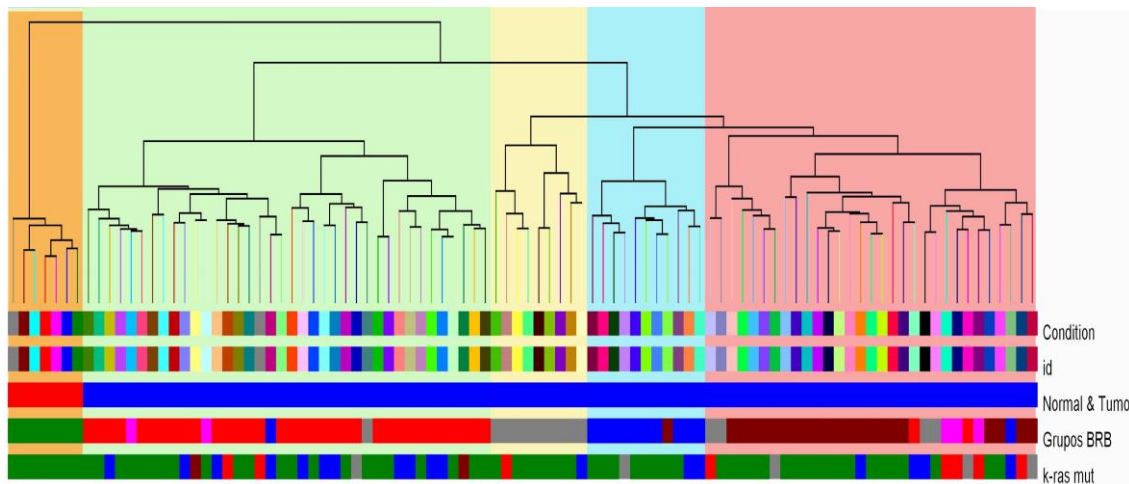


Tabla 4.2.Agrupación jerárquica de las muestras.

Fuente: elaboración propia

Los 14764 genes explican toda la variabilidad de cada una de las 96 muestras y establecen una ordenación en función de su variabilidad. Si al realizar una selección de genes, éstos son capaces de ordenar a las muestras de manera similar, significaría que se podría trabajar con este nuevo subconjunto de genes, puesto que nos estarían aportando la misma información que los datos originales de partida.

Como puede observarse en la tabla 4.2., al ordenar nuevamente las muestras únicamente con los genes cuya beta sea mayor a uno coma seis y en, al menos, más de 8 muestras el gen esté fuera del intervalo seleccionado, la ordenación se mantiene muy similar.



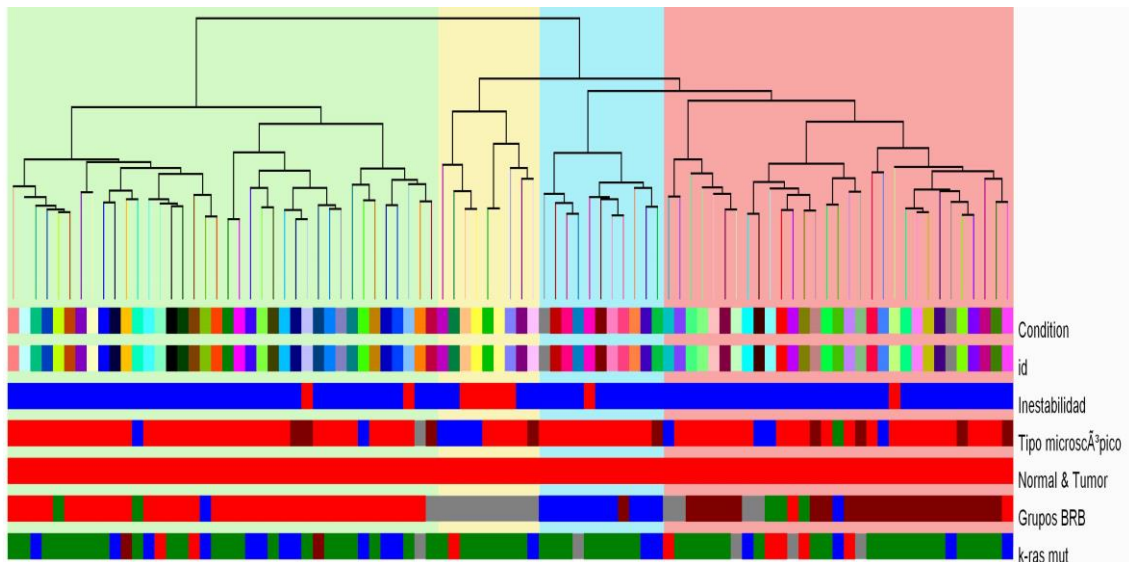


Imagen4.4: Gráfico del clustering efectuado.

Fuente: elaboración propia.

El clúster de color verde clasifica por completo al subgrupo 1 de cáncer colorrecta, el que presenta bajos niveles de estroma. El clúster de color azul clasifica al subtipo 2, cáncer colorrectal relacionado con las inmunoglobinas. El clúster de color rojo clasifica al subtipo 3, cáncer colorectal que presenta alto en de estroma.Finalmente, el clúster de color amarillo clasifica al subtipo 4, cáncer de colon inestable mucinoso

Por todo ello, se puede concluir que la técnica empleada en este trabajo que selecciona 521 genes de los 14764 iniciales y los clasifica en virtud del parámetro beta, es válida para las 96 muestras analizadas.

## 4.2. IMPLICACIONES

Se ha propuesto un marco teórico, novedoso en el campo de la oncología molecular, para la identificación de eventuales biomarcadores y genes diana. Este trabajo no pretende dar la solución única al problema de qué metodología debe seguirse para la identificación y selección de genes diferencialmente expresados. Únicamente, se desea aportar una nueva propuesta con base en los perfiles de expresión<sup>67</sup> obtenidos en función de las “betas” de los genes que clasifica a los tumores no por su histopatología, fundamento en el que se sustentan las principales clasificaciones tumorales como por ejemplo: Dukes<sup>68</sup>, Astler y Coller<sup>69</sup> o TNM, sino por la variabilidad en la expresión de los genes. Quizás, pueda contribuir, mediante esta nueva enfoque, en un futuro cercano a explicar las razones por las cuales un mismo tratamiento en pacientes oncológicos con el mismo tumor y estadio genera diferentes respuestas en los pacientes.

De validarse esta técnica en otros trabajos sería una herramienta más en la selección o filtrado de genes haciéndose extensible a otros tipos de tumores. Las posibles implicaciones que tendría una mejora en los métodos de identificación de genes en el campo sanitario son muchas, entre los que podemos destacar:

- Mejora en los tratamientos personalizados para los pacientes.

---

<sup>67</sup> Valor de la expresión de un gen que nos facilita el microarray.

<sup>68</sup> Publicada en 1932. Para más información véase el Anexo IV

<sup>69</sup> Publicada en 1954. Para más información véase el Anexo IV

- Avance en el diseño de nuevos fármacos mucho más específicos y con menos toxicidad que los existentes mejorando la calidad de vida de los pacientes con cáncer.
- Avance en la prevención del cáncer.
- Aumento de la esperanza de supervivencia de los pacientes oncológicos.
- Utilización de los datos facilitados tanto para la prevención como para el diagnóstico y pronóstico clínicos.

Otra mejora a tener en cuenta, fuera del ámbito clínico, sería el económico. El tratamiento oncológico, en estos momentos, es uno de los más caros. Si se logran desarrollar fármacos más específicos y se avanza en los tratamientos personalizados, el coste sanitario también disminuiría.

Tomando como punto de partida el estudio llevado a cabo en el Hospital Doce de Octubre de Madrid, los doctores Antonio Jimeno, de la Universidad Johns Hopkins (Estados Unidos), Hernán Cortés-Funes, del Doce de Octubre, y Ramón Colomer, del Instituto Catalán de Oncología (ICO), concluían que este coste rondaba los 30.000 euros por paciente, para una supervivencia media de 25 meses. No se incluyen en esta cuantía otros medicamentos de apoyo pero que, a menudo, suman cifras mayores incluso que la propia quimioterapia. Un hospital madrileño de tamaño medio, por ejemplo, puede gastar más de 50.000 euros mensuales en fármacos para combatir los vómitos (antieméticos) causados por la 'quimio', unos 35.000 euros en eritropoyetinas (que combaten

la anemia de estos pacientes) y casi 7.000 en factores de crecimiento medular, que fortalecen el sistema inmunitario.

Si se lograsen identificar genes diana se abriría la puerta para desarrollar fármacos más específicos a la vez que se avanzaría en los tratamientos personalizados. El coste sanitario derivado del tratamiento oncológico, lógicamente, disminuiría al tener los pacientes menos efectos secundarios ante tratamientos mucho más específicos, personalizados y efectivos.



### **4.3. RECOMENDACIONES**

El estudio realizado en esta Tesis Doctoral contempla 96 muestras: 89 tumorales y 7 de normales o de control. Sería deseable disponer de un mayor número de pacientes que participaran en el estudio. Ciertamente, que los datos de que disponen los equipos de investigación no suelen superar esta cifra. Hoy en día, gracias a la creación de las redes de bancos de tumores se puede tener acceso a un número de muestras sensiblemente superior.

Además, el avance en computación de los últimos años, nos permite disponer de computadores y aplicaciones informáticas capaces de analizar grandes volúmenes de información.

Ambas razones, la disponibilidad de bancos de tumores y la capacidad de computación actual, contribuyen a valorar la posibilidad de repetir la metodología desarrollada en este trabajo en un estudio más amplio. Las técnicas planteadas en esta Tesis no son exclusivas para el cáncer de colon sino que son, perfectamente, extensibles a cualquier otro tipo de tumor.

## BIBLIOGRAFÍA

Alexander, G., Sharpe, W. y Bailey, J (2003). *Fundamentos de Inversiones: Teoría y Práctica*. Mexico: Pearson.

Alizadeh, A, Eisen M., Davi, R., Mac, L. y Rosenwald A. “*Distinct Types of Diffuse Large b-cell Lymphoma Identified by Gene Expression Profiling*”. Nature. 2000 feb 3; 403 (6769):503.

Alonso, A. (2006). *Introducción al Uso del R en Econometría*. Madrid: Ediciones Escorialenses.

Ama,t D. y Puig, X. (1989). *Análisis Técnico Bursátil*. Barcelona: Gestión 2000.

Aragoneses, J.R. (1.986). *Análisis del Comportamiento de los Rendimientos bursátiles*. Madrid: Gestión Científica nº 3.

Aragoneses, J.R., y Blanco, C. (2000). *Valor en Riesgo: Aplicaciones a la Gestión Empresarial*. Madrid: Pirámide.

Ballart, L. (1988). *Proceso de Innovación en el Sistema Financiero Español*. Instituto de Empresa. Madrid.

Banz, R.W. (1.981). *The Relationship Between Return and Market Value of Common Stock*. Journal of Financial Economics nº 9. Marzo. Págs.: 3-18.

Bakay, M, Yi-Wen, C., Rehannah, Po Zhao, Kanneboyina N., Eric P. *Sources of Variability and Effects of Experimental Approach on Expression Profiling Data Interpretation*. Mml Bioinformatics, Biomed. Central.

Blanco, F., Fernando, M. y Martínez, F. (2005). *Dirección Financiera I: Selección de Inversiones*. Madrid: Pirámide.

Bob Ryan, Robert W. y Theobald M. (2005). *Metodología de la Investigación Financiera en Finanzas y Contabilidad*. Barcelona: Deusto.

Boldt, H.L. (1984). "Efficient Markets and the Professional Investors". *Financial Analysts Journal*, July-August 1984.

Black, F. y Scholes (1972) *The Valuation of Option Contracts and a Test of Market Efficiency*. *Journal of Finance*, Mayo 1972.

Bradley, E. Robert T., John D. y Virginia T. (2003). *Empirical Bayes Analysis of a Microarray Experiment*.

Bradley, Efron, Trevor Hastie, Iain Johnstone y Robert Tibshirani (2003). *Least Angle Regression*. *Biometrika Journal of Cancer*.

Brealey - Myers (2005). *Principios de Finanzas Corporativas*. Madrid: McGrawhill.

Bradley, S.P. y Kane, D.B. (1973) "Management of Commercial Bank Government Security Portfolios: An optimization Approach Under Uncertainty". *Journal of Bank Research*. Primavera.

Bradley, S.P. y Kane, D.B. (1975). *Management of Bank Portfolios*. John. Nueva York: Wiley & Sons.

Bradley, R.A. (1990) "Portfolio Theory versus Portfolio Practice". *Journal of Portfolio Management*. Verano. Págs.: 6-10

Bruce, A, Alexander J., Julian, L., Martin R, K, Meter W. (2003). *Biología Molecular de la Célula*. Barcelona: Ediciones Omega.

CFA Series - *Investment Analysis and Portfolio Management* (Reilly Brown) 7th ed.

Codina J. (1997). *Manual de Análisis Técnico*. Madrid: Inversor Ediciones.

Colby W. y Meyers A. (1992). *Enciclopedia Mundial de los Osciladores Técnicos del Mercado*. Madrid: Gesmovasa.

Córdoba Bueno Miguel (2000). *La Práctica de los Mercados Financieros*. Madrid: Dykinson.

Cruz Hernández J.J. (2004). *Lecciones de Oncología Clínica*. Madrid: Nova Sidonia.

Dai H., Van't Veer L., Lamb J., He Yd, Mao M., Fine Bm, et all. "A Cell Proliferation Signature is a Marker of Extremely Poor Outcome in a Subpopulation of Breast Cancer Patients". *Cancer res.* 2005 may 15;65 (10):4059-66.

Dattatreya E., Venkatesh E. S.; Venkatesh, Vi. (1995). *Swaps Sobre Tipos de Interés y Sobre Divisas*. Madrid: Gesmovasa.

Díez de Castro L. y Mascareñas J. (1994). *Ingeniería Financiera. La Gestión de los Mercados Internacionales*. Madrid: McGraw Hill.

Díaz E. y García-Conde J. (2000). *Oncología Clínica Básica*. Madrid: Arán.

Driver, C. y Moreton, D. (1993). *Inversión, Expectativas e Incertidumbre*. Madrid: Celeste Ediciones.

Ebert BL y Golub TR. "Genomic Approaches to Hematologic Malignancies". *Blood*. 2004 aug 15;104(4):923-32.

Edwards D.; Magee J.(1995). *Análisis Técnico de las Tendencias de los Valores*. Madrid: Gesmovasa.

Edwards, R. y Magee, J. (1984) *Technical Analysis of Stock Trends*. Boston: John Magee Inc.

Elton, E.J. y Gruner M.J. (1984) *Modern Portfolio Theory and Investment Analysis*. Nueva York: John Wiley & Sons.

Enatzari, J y Loamnids P. (2003) *Predictive Ability of DNA Microarrays for Cancer Aoutcomes and Correlates: an Empirical Assessment*. Lancet 2003; 362: 1439-44.

Fama E.F. (1965) “*Random Walks in Stock Markets*”. Financial Analyst Journal. Sept-oct. Págs.: 34-105.

Fama E.F. (1970) “*Efficient Capital Market: A Review of Theorie and Empirical Work*”. Jonrnal of Finance nº 25. Mayo. Págs.: 383-417.

Fama E.F. (1965) “*The Behavior of Stock Market Prices*”. Journal of Business, nº 38. Enero. Págs.: 34-105.

Fama E.F. y FRENCH K. (1992) “*The Cross Section of Expected Stocks Returns*”. Journal of Finance vol. 47. nº 2. Junio. Págs.: 427-465.

Fama E.F. (1968) “*Risk, Return and Equilibrium: Some Clarifying Commnets*”. *Journal of Finance*. Marzo. Págs.: 29-40.

Fama E.F. (1984) “*Term Premium in Bond Returns*”. Economics, nº 3, Diciembre, Págs 529-546.

Fama E.F. y Macbeth, J.D. (1973) "*Risk, Return and Tests*. Journal of Political Economy. Mayo-junio. Págs 561-75.

Fama E.F., Jensen, M.C. y ROLL, R. (1969) "*The Adjustment of New Information*". International Economic Review. Enero. Págs.: 1-21.

Feria J.M. (2005) *El Riesgo de Mercado: Su medición y Control*. Madrid: Delta.

Fernández Blanco, M y otros (1992) *Dirección Financiera de la empresa*. Madrid: Pirámide.

Fernando Bolado, M (1982) *Aplicación del Modelo de Equilibrio al Mercado de Capitales*. Tesis Doctoral. Universitat de València.

Fernández M, Gómez AR. Piñol JA y Reig Pérez A (2005) *Teoría de la Financiación I: Modelos CAPM, APT y Aplicaciones*. Madrid: Ediciones Pirámide.

Fernández, P. (1998) "*EVA, Beneficio Económico y Creación de Valor*". Bolsa de Madrid; Num 70; Oct 1998.

Fernández, P. (1989) *Bonos Convertibles en España*. Barcelona: IESE.

Fisher, D.E. y Jordan, D.J. (1975) *Security Analysis and Portfolio Management*. Prentice Hall. 5ª ed.

Fisher, J. (1965) "*The Theory of Interest*". Augustus M. Kelley, Publishers, Nueva York. Reimpreso a partir de la edición de 1930.

Fisher, L. y Weil, R.L. (1971). "*Coping with the Risk of Interest Rate Fluctuations: Returns to Bondholders from Naive and Optimal Strategies*". Journal of Business. Octubre. Págs.: 408-431.

Friedman, M. (1983). *Libertad de Elegir*. Barcelona: Orbis.

FROST y PRECHTER. (1989) *El Principio de la Onda de Elliot*. Madrid: Gesmovasa.

Galán F. (2004) *Riesgo, Rentabilidad y Eficiencia de Carteras de Valores*. Colección Etea.

Galbraith, J.K. (1983) *El dinero*. Orbis.

García-Gutiérrez, C., Mascareñas, J. y Pérez, E. (1998) *Casos Prácticos de Inversión y Financiación en la Empresa*. Madrid: Pirámide.

García A. (2004) *Métodos Avanzados de Estadística Aplicada, Técnicas Avanzadas*. Madrid: Educación Permanente Programa de Enseñanza Abierta.

García A. (2004) *Métodos Robustos y de Muestreo*. Madrid: Educación Permanente Programa de Enseñanza Abierta.

García J. (2008) *Oncogenes y Genes Supresores de Cáncer*. Madrid: Arán.

García J., Martínez E., Redondo R., del Campo C. (2002) *Métodos de Decisión*. Madrid: Prentice Hall.

García P. (2008) *Nuevas Dianas Terapéuticas en Cáncer Colorrectal*. Madrid: Arán.

Gardner, R. (1996) *Juegos para Empresarios y Economistas*. Barcelona: Antoni Bosch, editor, S.A.

Glas AM, Kersten MJ, Delahaye L, Witteveen A, Kibbelaar R,. “*Gene Expression Profiling in Follicular Lymphoma to Assess Clinical Aggressiveness and to Guide the Choice of Treatment*”. Blood. 2005 jan 1;105(1):301-7.

Geoffiey FI. Moore. (1980) “*Business Cycles, Inflation, and Forecasting*”. National Bureau of Economic Research. Studies on Business Cycles, nº. 24, 1980.

Gómez-Bezares, F. (1992) *25 Años de Contratación del CAPM*. Actualidad Financiera, nº 32. Septiembre. Págs.: F563-F588.

Gómez-Bezares, F. (2006) *Gestión de Carteras: Eficiencia, Teoría de carteras, CAPM, APT*. Madrid: Desclee de Brouwer.

Harrington, DR. (1987) *Modern Portfolio Theory, The Capital Asset Pricing Model and Arbitrage Pricing Theory*. New Jersey: Prentice- Hall.

Hicks, I.R. (1939) *Value and Capital*. New York: Oxford University Press.

Horra J. (2000) *Estadística Aplicada*. Madrid.

Instituto Español de Analistas Financieros (1996) *Curso de Bolsa y Mercados Financieros*. Madrid: Ariel.

James T. (1992) *Opciones Sobre Futuros*. Madrid: Gesmovasa.

Joern T., Sebastian S., Matthias H., Benjamin G. y Stefan R. (2005) “*Microarray Chromosome Analysis Tool*”. Bioinformatics Applications Note. vol. 21 no. 9 2005, pages 2112-2113 doi: 10.1093/bioinformatics/bt/183.

Keynes, J.M. (1981) *Teoría general de la Ocupación, el Interés y el Dinero*, Madrid: FCE



Lintner, J. (1.965) “*Security Prices Risk and Maximal Gains from Diversification*”. *Journal of Finance*. Diciembre. Págs.: 587-615.

Lintner, J. (1.965) “*The Valuation of Risk Assets and the Selection of Risk Investments in Stock Portfolios and Capital Budgets*”. *The Review of Economic and Statistics* vol. XLVII. nº 1. Febrero. Págs.: 13-37.

López J. y Rojo J. (2004) *Los Mercados de valores*. Madrid: Pirámide.

Marchal JA., Gaforio JJ., y Aránega A. (2006). *Avances en Oncología Básica y Aplicada*. Granada: Universidad de Granada.

Markowitz, H. (1952) “*Portfolio selection*”. *Journal of Finance* 7. 1952. Marzo. Págs.: 77 a 91.

Markowitz, H. (1959) “*Portfolio Selection: Efficient Diversification of Investments*”. 1959 Nueva York: John Wiley & Sons.

Martín J. (1989) *Análisis Técnico Explicado I*. Madrid: Gesmovasa.

Martín J. (1994) *Mercados Derivados Sobre Índices Bursátiles*. Madrid: Gesmovasa.

Mascareñas, J. (2004) *El Riesgo en la Empresa. Tipología, Análisis y Valoración*. Madrid: Pirámide.

Mascareñas, J. (1991) *La Estructura Temporal de los Tipos de Interés*. *Actualidad Financiera* nº 18. Abril-Mayo. Págs.: 213.

Mascareñas, J. (1996) *La Gestión de las Carteras de Renta Fija*. Madrid: Ed. Abacus. Cap. 5 En prensa.

Mascareñas, J. (1991) *La Gestión Financiera de las Carteras de Renta Fija (III): Gestión Activa y Pasiva*. Actualidad Financiera nº 24. Junio. Págs.: F-444-445.

Mascareñas, J. (1991) “*La Estructura Temporal de los Tipos de Interés*”. Actualidad Financiera nº 18. Abril-Mayo. Págs.: 201-226.

Mascareñas J. (1999) *Innovación Financiera: Aplicaciones para la Gestión Empresarial*. Madrid: McGraw Hill.

Mascareñas J., Lamothe P., López F. y Luna W. (2004) *Opciones Reales y Valoración de Activos*. Madrid: Prentice Hall.

Mei-Ling L., Frank C., kuo G., y Jeffrey S. (2000) “*Important of Replication in Mycroarray Gene Expresión Studies: Statistical Method and Evidence from Repetitive cDNA Hybridation*”.

Meyerhardt J. y Saunders M. (2007) *Cáncer Colorrectal*. Madrid: Elsevier Mosby.

Modigliani, F. y Miller, M (1958) “*The cost of Corporation Finance and the Theory of Investment*”. The American Economics Review. Junio. Págs.: 261-297.

Modigliani, F. y Miller, M (1963) “*Corporate Income, Taxes and the Cost of Capital: a Correction*”. The American Economics Review. Junio. Págs.: 433-443.

Modigliani, F. y Modigliani L. (1997) “*Risk Adjusted Performance*”. Journal of Portfolio Management. Invierno. Págs.: 45-54.

Muñoz A. (1997) *Cáncer: Genes y Nuevas Terapias*. Madrid: Hélice.

Murphy JJ. (2000) *Análisis Técnico de los Mercados Financieros*. Barcelona: Ediciones Gestión 2000.

Novales A. (2000) *Econometría 2ª ed*. Madrid: McGraw Hill.

Padillo N. (2004) *Análisis Multicriterio Aplicado a la Gestión de Carteras*. Universidad de Huelva publicaciones.

Palacios R. y Rofman R (2004) *Gene Expression Test for Breast Cancer's Future*. Science vol. 303; 19 March 2004.

Quackenbush J. (2005) *Computational Analysis of Microarrays Data*. The Institute of Genomic Research, 9712 Medical Center.

Quesada A.R., Medina M.A. y Sánchez F. (1999) *Biología Molecular del Cáncer: Fundamentos y Perspectivas*. Málaga. Universidad de Málaga.

Rodríguez A., Martínez N., Piris MA. (2004) "Variability in Degree of Expression of Phosphorylated in Chronic Lymphocytic Cases with Nodal". *Clinical Cancer Research* 6796 vol.10, 6796-6800. October 15, 2004.

Rosell R., Abad A., Monzó M y Barnadas A. (2006) *Oncología Clínica y Molecular*. Madrid: Arán.

Rubio R. (2001) *Manual de Inversión en Bolsa*. Madrid: Inversor Ediciones S.L.

Sanz AI., Sáez AJ., Rovero C., Piris MA. (2004) *Building an Outcome Predictor Model for Diffuse Large b-cell Lymphoma*. *American Journal of Pathology* vol. 164, no.2. February 2004.

Sánchez, J.L. *Curso de Bolsa y Mercados Financieros*. Ariel.

Schumpeter, J.A. (1982) *Historia del análisis económico*. Barcelona: Miel.

Sewart, T.H. (1988). *Cómo Pueden Proporcionarle Dinero los Gráficos*. Madrid: Gesmovasa.

Sharpe, W.F. (1964) "*Capital Asset Prices: A Theory of Market Conditions of Risk*". *Journal of Finance* nº 19. Septiembre. Pág.: 425-442.

Sharpe, W.F. y Gordon J.A. (1989) *Fundamentals of Investments*. Prencice-Hall. Englewood Cliffs (NJ).

Sharpe, W.F. (1963) "*A Simplified Model for Portfolio*". *Science* vol. IX nº 2 Enero. Págs.: 277-293.

Sharpe, W.F. (1966) *Mutual Fund Performance*. *Journal Equilibrium under Analysis*. Management of Bussines nº 39. Págs.:93-130.

Sharpe, W.F. (2007) *Inversores y Mercados*. Barcelona: Deusto.

Sharpe, W.F. (1964) "*Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk*". *The Journal of Finance*, vol XIX. Nº3. Septiembre 1964.

Sharpe, W.F. (1970) *Portiolio Theory and Capital Markets*. Nueva York: McGraw-Hill.

Sharpe, W.F. (1975) *Teoría de Cartera y del Mercado de Capitales*. Barcelona: Deusto.

Stefan, M., Serge K., Catherine H. (2005) "*Prediction of Cancer Outcome with Microarrays: a Multiple Random Validation System*". *The Lancet*. feb 5-feb 11, 2005; 356,9458;proquest medical library.

Suarez, A. (1996) *Decisiones Óptimas de Inversión y Financiación en la Empresa*. Madrid: Pirámide.

Symon, R. (2003) “*Diagnostic and Prognostic Prediction Using Gene Expression*”. Brithis Journal of Cancer.

Tabernero, J. (2008) *Cáncer Colorrectal*. Madrid: Arán.

Taylor, F. (1996) “*Mastering Derivatives Markets*”. FT Pitman Publishing; 1996.

Tobin J. (1958) “*Liquidity Preference as Behavior Towards Risk*”. Review of Economics Studies, vol 26. Febrero 1958. Págs.: 65-86.

Treynor, J. (1965) “*How to Rate Management of Investment Funds*”. Harvard Business Review. Vol 43 nº 1. Enero-Febrero. Págs.: 63-75.

Treynor, J. y Mazuy F. (1966) “*Can Mutual Funds Outguess the Markets?*”. Harvard Business Review. Vol 44 nº 4. Págs.: 131-136.

Uriel E. y Aldas J. (2005) *Análisis Multivariante Aplicaciones al Marketing, Investigación de Mercados, Economía, Dirección de Empresa y Turismo*. Madrid: Thomson.

Watson J., Baker T., Bell S., Gann A., Levine M. y Losick R. (2006) *Biología Molecular del Gen*. Madrid: Panamericana.

## **ANEXO I. ADN, ARN Y EL CICLO CELULAR**

### **1.1. ADN.**

#### **1.1.1. ESTRUCTURA**

#### **1.1.2. BASES NITROGENADAS Y ELEMENTOS**

#### **1.1.3. PAPEL DE LA SECUENCIA**

#### **1.1.4. REPLICACIÓN DEL ADN**

#### **1.1.5. TRADUCCIÓN DEL ADN**

### **1.2. ARN**

#### **1.2.1. TIPOS DE ARN**

### **1.3. GENES, CROMOSOMAS, CROMATINA Y NUCLEOSOMA**

### **1.4. REGULACIÓN DEL CICLO CELULAR**

## ANEXO I. ADN, ARN Y EL CICLO CELULAR

### 1.1. ADN<sup>70</sup>.

ADN es el acrónimo de Ácido Desoxirribonucleico (en inglés, DNA: Deoxyribonucleic Acid). El ADN no sólo lleva la información esencial, sino que además es capaz de autorreplicarse, es decir, de dar lugar a nuevas copias de sí mismo, de manera que cuando una célula se divide para originar dos células hijas ambas adquieren copias iguales de ADN. Esta capacidad de asegurar la transmisión de la especificidad de las proteínas celulares es esencial para la vida y constituye la base de la genética.

Los genes son parte de la molécula de ADN. Los genes residen en los cromosomas que están localizados en el núcleo de la célula y son los transportadores de la herencia. Cada una de las aproximadamente  $10^{13}$  a  $10^{14}$  células (diez a cien billones) del cuerpo humano tiene el mismo número de cromosomas (46) y por tanto de genes (unos 40500). El contenido total del ADN de una célula o una especie se llama genoma.

Constituye el principal componente del material genético de la inmensa mayoría de los organismos, junto con el ARN. Es el componente químico primario de los cromosomas y el material en el que los genes están codificados.

---

<sup>70</sup> Este epígrafe ha sido elaborado según el manual *Biología Molecular del Gen* 5ª ed (pp. 107-132). Publicado por la Editorial Médica Panamericana (2006).

Su función es codificar las instrucciones esenciales para fabricar un ser vivo idéntico a aquel del que proviene (o casi similar, en el caso de mezclarse con otra cadena como es el caso de la reproducción sexual o de sufrir mutaciones).

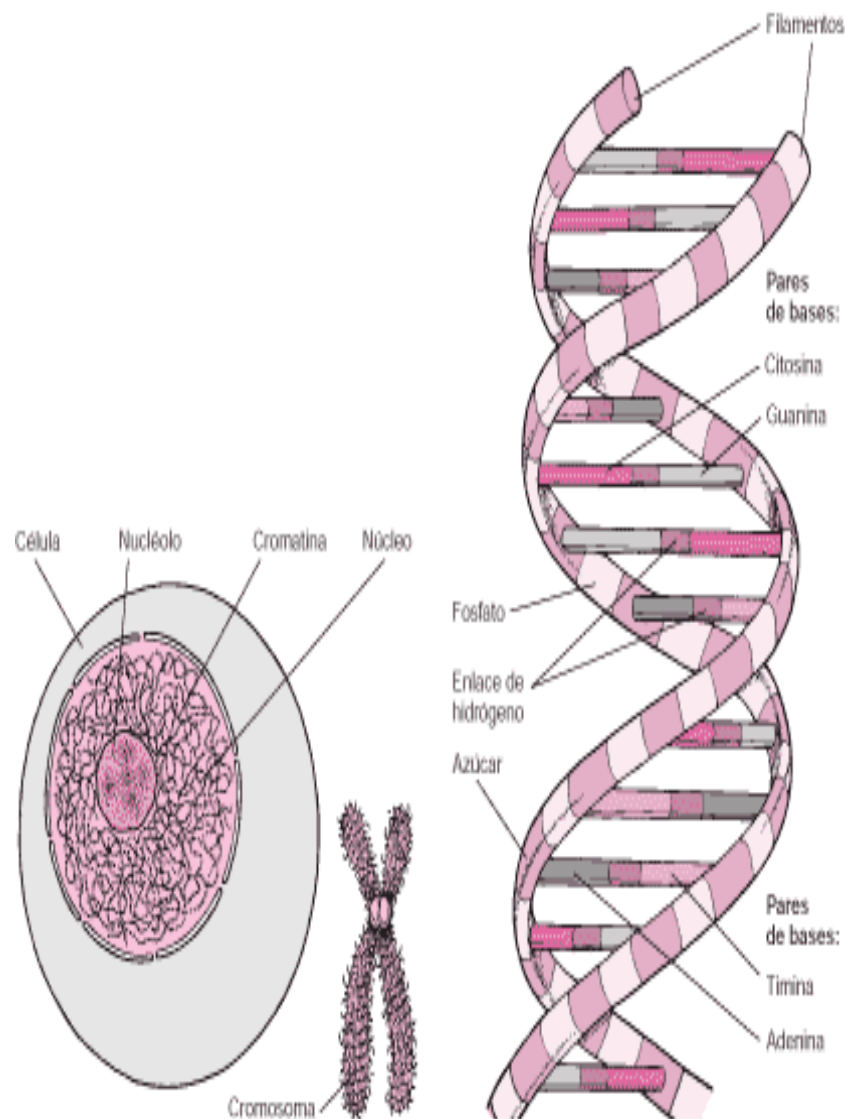


Imagen 5.

Fuente:[http://a248.e.akamai.net/7/248/847/20060531125413/www.msd.es/publicaciones/mmerck\\_hogar/seccion\\_01/images/seccion\\_01\\_03.gif](http://a248.e.akamai.net/7/248/847/20060531125413/www.msd.es/publicaciones/mmerck_hogar/seccion_01/images/seccion_01_03.gif)



### 1.1.1. Estructura

La característica más importante del ADN es que está compuesta por dos cadenas de polinucleótidos enroscadas una alrededor de la otra en la forma de una hélice doble. La columna de cada cadena la hélice de ADN está formada por monosacáridos y fosfatos; las bases se proyectan hacia el interior. La estructura del ADN es una pareja de largas cadenas de nucleótidos. La estructura de doble hélice del ADN fue descubierta en 1953 por James Watson y Francis.

Empecemos por considerar la naturaleza del nucleótido, la subunidad fundamental del ADN. Éste presenta un fosfato unido a un monosacárido (llamado 2'-desoxirribibosa), al que se adhiere una base. El monosacárido tiene las estructuras que aparecen en la siguiente figura:

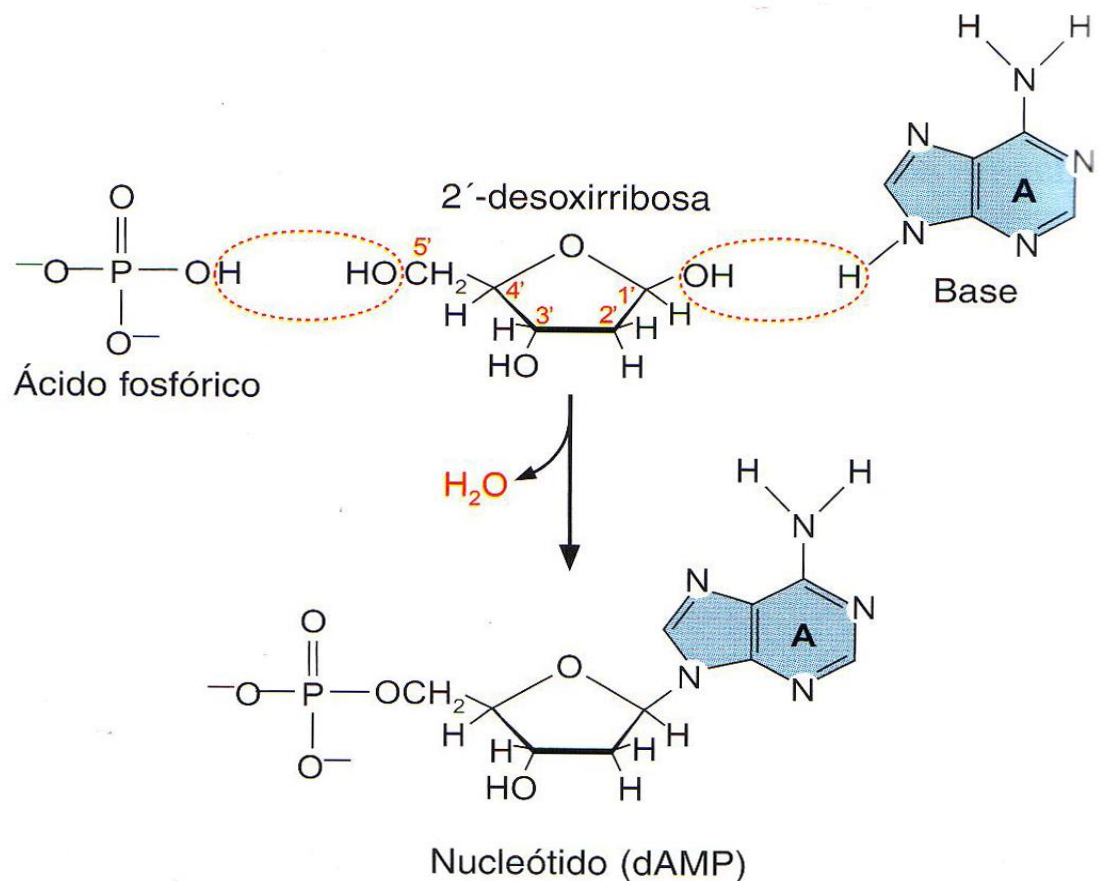


Imagen 6.

Fuente: Biología molecular del gen. 5ª edición, pag. 108

El monosacárido y la base solos reciben el nombre de nucleósido. La adición de un fosfato (o más de éstos) a un nucleósido forman un nucleótido. En consecuencia, mediante la formación de un enlace glucosídico entre la base y el monosacárido y mediante la formación de un enlace fosfoéster entre el monosacárido, y el ácido fosfórico se crea un nucleótido.

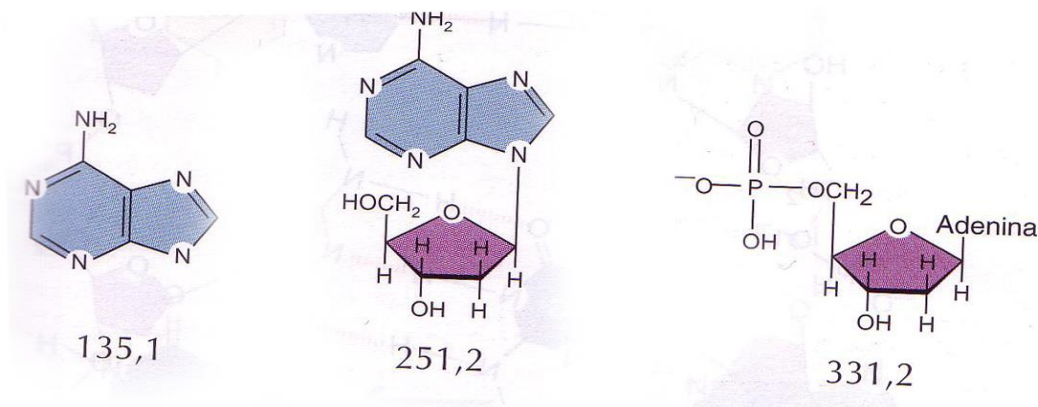


Imagen 7.

Fuente: Biología molecular del gen. 5ª edición, pag. 109.

A su vez, los nucleótidos se unen entre sí en cadenas de polinucleótidos por medio del hidroxilo 3' de la 2'-desoxirribosa de un nucleótido y el fosfato adherido al hidroxilo 5' de otro nucleótido.

Los nucleótidos están diferenciados por sus bases nitrogenadas divididas en dos grupos: dos purínicas (o púricas) denominadas adenina (A) y guanina (G) y dos pirimidínicas (o pirimídicas) denominadas timina (T) y citosina (C). El rasgo fundamental es que cada base nitrogenada de una hebra "casa" con la base de la otra, en el sentido de que la adenina siempre se enfrenta a la timina (lo que se denomina A-T) y la guanina siempre a la citosina (G-C). La adenina se une a la timina mediante dos puentes de hidrógeno, mientras que la guanina y la citosina lo hacen mediante tres puentes de hidrógeno; de ahí que una cadena de ADN que posea un mayor número de parejas de C-G sea más estable.

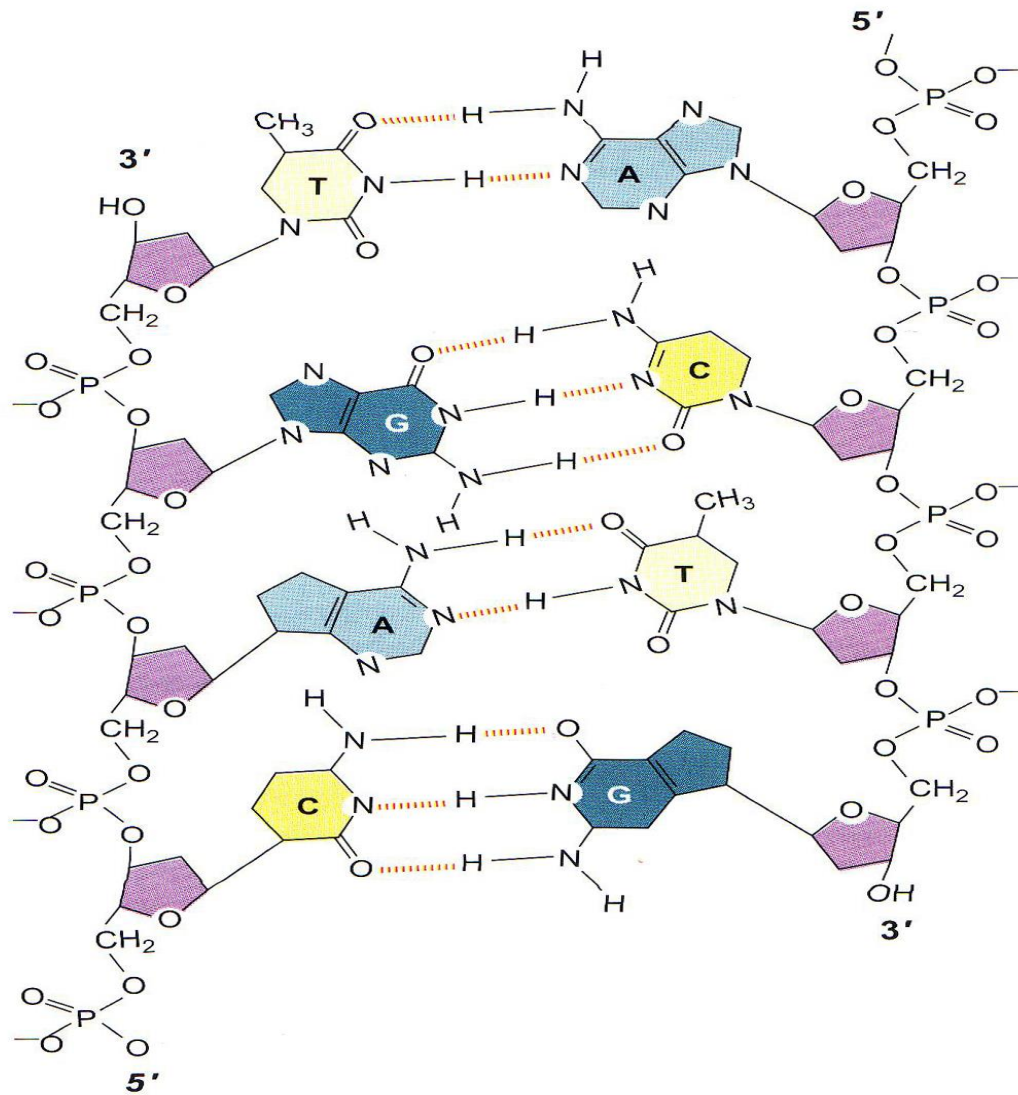


Imagen 8.

Fuente: Biología molecular del gen. 5ª edición, pag. 110.

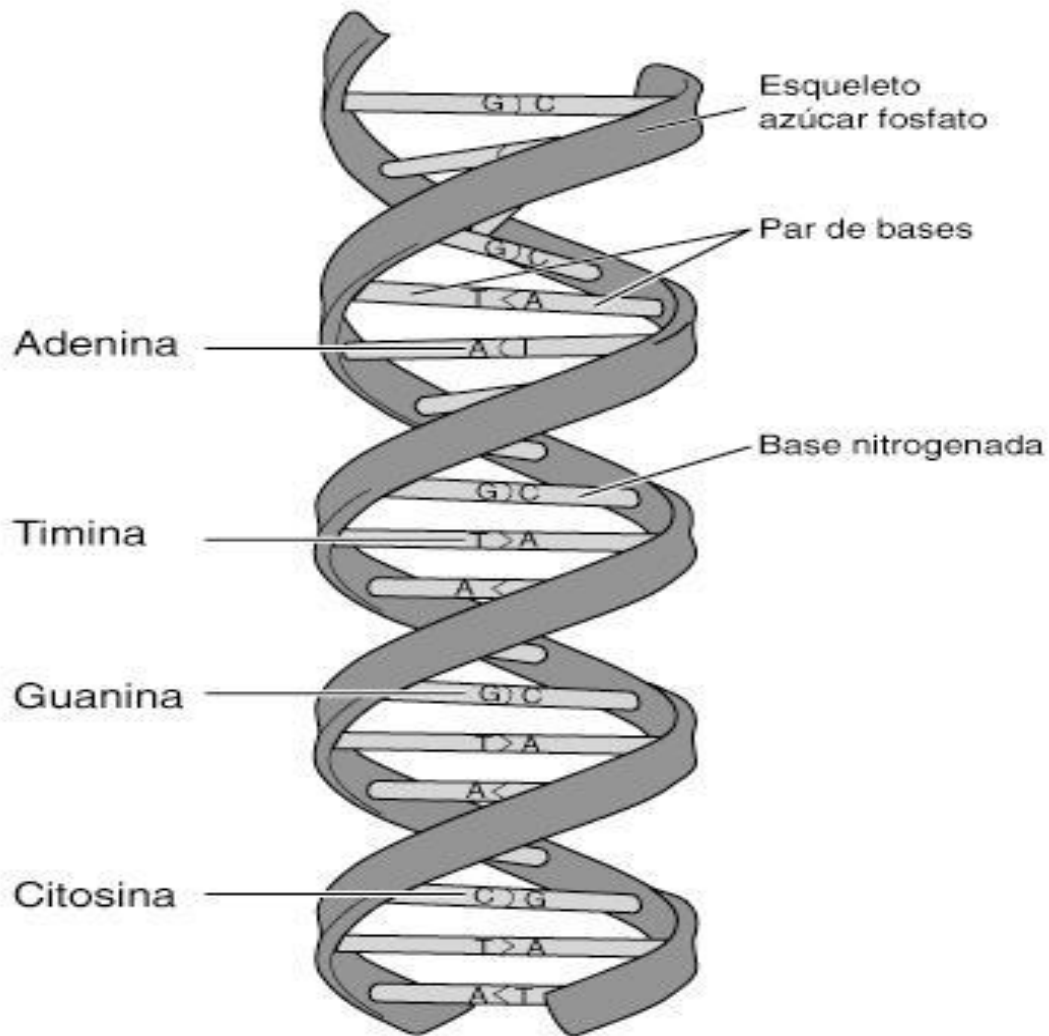


Imagen 9.

Fuente: <http://es.wikipedia.org/wiki/Imagen:ADN.jpg>

### 1.1.2. Bases nitrogenadas y elementos

Existen cuatro bases: dos púricas denominadas adenina (A) y guanina (G), y dos pirimidínicas denominadas citosina (C) y timina (T). La estructura del ADN es una pareja de largas cadenas de nucleótidos. El complemento es el siguiente:

La Adenina (A) siempre se empareja con la Timina (T) y la Citosina (C)  
siempre se empareja con la Guanina (G).

Adenina (A) con Timina (T)

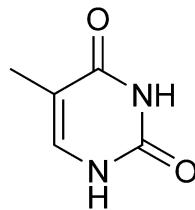
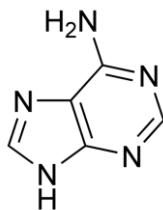
Citosina (C) con Guanina (G)

A - T

C – G

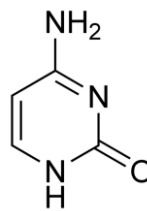
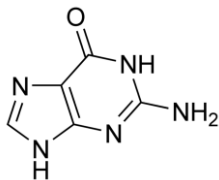
La adenina

Timina



Guanina

Citosina



### 1.1.3. Papel de la secuencia

En un gen, la secuencia de los nucleótidos a lo largo de una hebra de ADN se transcribe a un ARN mensajero (ARNm) y esta secuencia a su vez se traduce a una proteína que un organismo es capaz de sintetizar o "expresar" en uno o varios momentos de su vida, usando la información de dicha secuencia.

La relación entre la secuencia de nucleótidos y la secuencia de aminoácidos de la proteína viene determinada por el código genético, que se utiliza durante el proceso de traducción o síntesis de proteínas. La unidad codificadora del código genético es un grupo de tres nucleótidos (triplete), representado por las tres letras iniciales de las bases nitrogenadas (por ej., ACT, CAG, TTT). Cuando estos tripletes están en el ARN mensajero se les llama codones. En el ribosoma cada codón del ARN mensajero interacciona con una molécula de ARN de transferencia (ARNt) que contenga el triplete complementario (denominado anticodón). Cada ARNt porta el aminoácido correspondiente al codón de acuerdo con el código genético, de modo que el ribosoma va uniendo los aminoácidos para formar una nueva proteína de acuerdo con las "instrucciones" de la secuencia del ARNm. Existen 64 codones posibles, por lo cual corresponde más de uno para cada aminoácido; algunos codones indican la terminación de la síntesis, el fin de la secuencia codificante; estos codones de terminación o codones de parada son UAA, UGA y UAG (en inglés, nonsense codons o stop codons).

#### 1.4. Replicación del ADN

Una propiedad esencial del material genético es su capacidad para hacer copias exactas de sí mismo, para lo cual cada una de las ramas de la cadena de ADN actúa como molde o guía, dirigiendo la síntesis de una nueva cadena complementaria a lo largo de su longitud, utilizando las materias primas de la célula. A medida que cada una de las ramas de la cadena originaria se

separan (rompiendo los puentes de hidrógeno entre sus bases nitrogenadas), cada una atrae nucleótidos complementarios (libres y disponibles en la célula), formando una nueva cadena. Este proceso ocurre una sola vez en cada generación celular, durante el segundo momento de la interfase y diferentes enzimas participan catalizando cada paso particular del proceso.

La iniciación de la replicación del ADN comienza siempre con una secuencia específica de nucleótidos conocida como el origen de replicación. Requiere proteínas iniciadoras especiales y además enzimas conocidas como helicasas, que rompen los puentes de hidrógeno abriendo la hélice, formándose las horquillas de replicación, una a cada lado de la burbuja a que da lugar la separación de las ramas del ADN. Una vez abierta la cadena de ADN, proteínas adicionales (conocidas como proteínas de unión a cadena simple o topoisomerasas) se unen a las cadenas individuales del ADN manteniéndolas separadas y evitando que se retuerzan. En el siguiente paso, las enzimas llamadas ADN polimerasa catalizan la síntesis real de las nuevas cadenas, añadiendo nucleótidos sobre el molde, las que se dan bidireccionalmente desde cada una de las horquillas que se replican en sentido opuesto dentro de cada burbuja, cuando éstas se encuentran y se fusionan todo el cromosoma ha quedado replicado longitudinalmente.

Para que el ADN polimerasa comience su tarea debe estar presente un cebador -molécula formada por nucleótidos de RNA catalizados por ARN primasas- que determina el punto por donde el ADN polimerasa comienza a



añadir nucleótidos, continuando por la cadena de ADN de molde en la dirección 5' a 3'. Debido a esta unidireccionalidad del ADN polimerasa, la replicación es continua en una de las ramas (cadena adelantada), mientras que en su antiparalela (cadena retrasada) es discontinua, fragmentada (siempre 5' a 3'); en ésta, cuando un ADN polimerasa hace contacto con el extremo de otro fragmento Okazaki el cebador de éste es eliminado y otra enzima, el ADN ligasa, conecta los segmentos de ADN recién sintetizado, catalizando las reacciones de condensación que unen los grupos fosfato y azúcar de los nucleótidos contiguos.

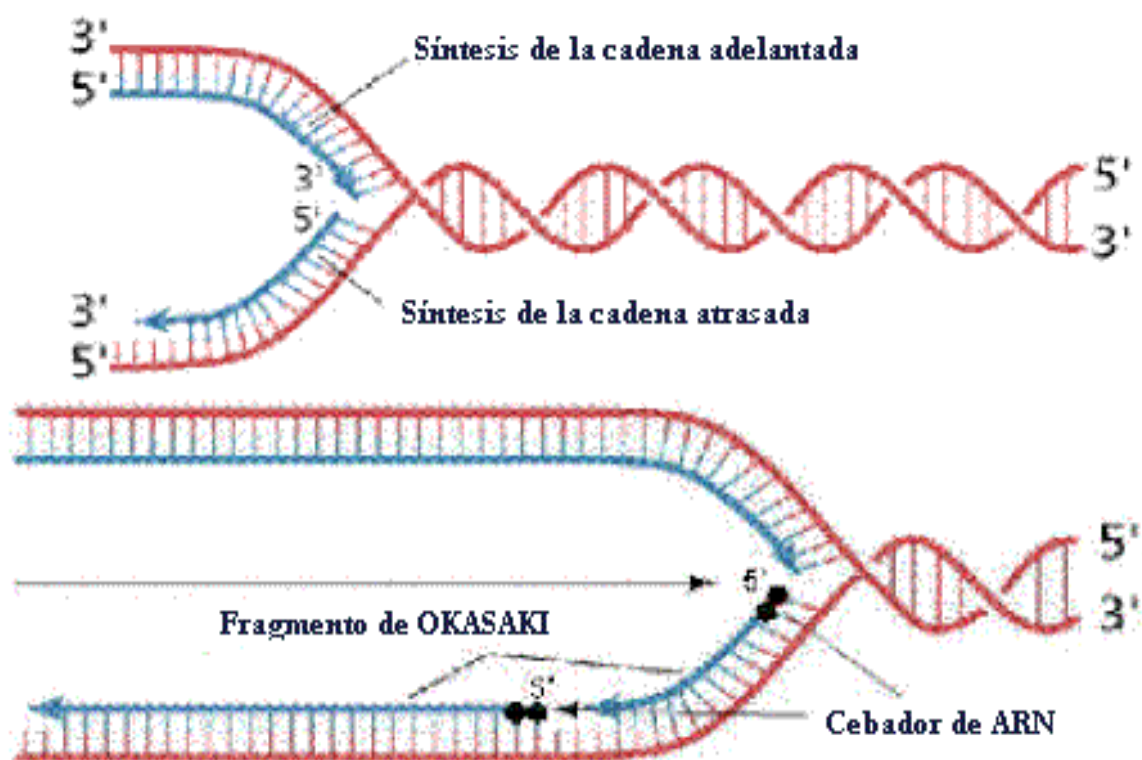


Imagen 11.

Fuente: <http://www.prodiversitas.bioetica.org/nota66-3.htm>.

## 1.5. Traducción del ADN

En un gen, la secuencia de los nucleótidos a lo largo de una hebra de ADN se transcribe a un ARN mensajero (ARNm) y esta secuencia a su vez se traduce a una proteína que un organismo es capaz de sintetizar o "expresar" en uno o varios momentos de su vida, usando la información de dicha secuencia.

La relación entre la secuencia de nucleótidos y la secuencia de aminoácidos de la proteína viene determinada por el código genético, que se utiliza durante el proceso de traducción o síntesis de proteínas. La unidad codificadora del código genético es un grupo de tres nucleótidos (triplete), representado por las tres letras iniciales de las bases nitrogenadas (por ej., ACT, CAG, TTT). Cuando estos tripletes están en el ARN mensajero se les llama codones.

En el ribosoma cada codón del ARN mensajero interacciona con una molécula de ARN de transferencia (ARNt) que contenga el triplete complementario (denominado anticodón). Cada ARNt porta el aminoácido correspondiente al codón de acuerdo con el código genético, de modo que el ribosoma va uniendo los aminoácidos para formar una nueva proteína de acuerdo con las "instrucciones" de la secuencia del ARNm. Existen 64 codones posibles, por lo cual corresponde más de uno para cada aminoácido; algunos codones indican la terminación de la síntesis, el fin de la secuencia codificante; estos codones de terminación o codones de parada son UAA, UGA y UAG (en inglés, nonsense codons o stop codons).

En muchas especies de organismos, sólo una pequeña fracción del total de la secuencia del genoma codifica proteínas; por ejemplo, sólo un 3% del genoma humano consiste en exones que codifican proteínas. La función del resto por ahora sólo es especulación, es conocido que algunas secuencias tienen afinidad hacia proteínas especiales que tienen la capacidad de unirse al ADN que tienen un papel importante en el control de los mecanismos de transcripción y replicación. Estas secuencias se llaman frecuentemente secuencias reguladoras, y los investigadores asumen que sólo se ha identificado una pequeña fracción de las que realmente existen. El llamado ADN basura representa secuencias que no parecen contener genes o tener alguna función; la presencia de tanto ADN no codificante en genomas eucarióticos y las diferencias en tamaño del genoma representan un misterio que es conocido como el enigma del valor de C.

Algunas secuencias de ADN desempeñan un papel estructural en los cromosomas: los telómeros y centrómeros contienen pocos o ningún gen codificante de proteínas, pero son importantes para estabilizar la estructura de los cromosomas.

## 1.2. ARN<sup>71</sup>

El ácido ribonucleico (ARN o RNA) es un ácido nucleico, polímero lineal de nucleótidos formando una larga cadena. El eje de la cadena lo forman grupos fosfato y azúcares ribosa de forma alternativa, del que toma su nombre. Los nucleótidos del ARN contienen el azúcar ribosa y entre sus bases nitrogenadas al Uracilo, a diferencia del ácido desoxirribonucleico (ADN) cuyo azúcar es una desoxirribosa y contiene a la Timina en vez del Uracilo. La función principal del ARN es servir como intermediario de la información que lleva el ADN en forma de genes y la proteína final codificada por esos genes.

En las células eucariotas el ADN se encuentra encerrado en el núcleo. La síntesis de ADN se hace en el núcleo, así como también la síntesis de ARN, pero la síntesis de proteínas ocurre en el citoplasma. El mecanismo por el cual la información se trasvasa desde el núcleo celular al citoplasma es mediante la transcripción del ARN a partir del ADN y de la traducción de proteínas a partir de ARN.

### 1.2.1. Tipos de ARN

ARN es el acrónimo de Ácido Ribonucleico. Existen distintos tipos de ARN:

---

<sup>71</sup> Este epígrafe ha sido elaborado según el manual *Biología Molecular del Gen* 5ª ed (pp. 133-140). Publicado por la Editorial Médica Panamericana (2006).

El ARN mensajero (ARNm) sufre diferentes fases durante su existencia que suele ser generalmente breve. El ARN mensajero se sintetiza en el núcleo celular en eucariotas mediante el proceso llamado transcripción del ADN.

Los ARN Transferentes (ARNt) reconocen los ARNm y transfieren un aminoácido determinado a la cadena de proteína que se está sintetizando. Según la información del ARNm, los ARNt sitúan los distintos aminoácidos en el lugar adecuado para sintetizar una cadena polipeptídica.

Un ARNt está formado por entre 73 y 90 nucleótidos.

El ARN nucleolar (ARNn) es una pequeña molécula de ácido ribonucleico, sintetizado y localizado en el nucléolo de las células eucariotas, a partir de la transcripción del ADN, formado por una corta secuencia de entre 100 a 300 nucleótidos, y que es precursor e indispensable para la síntesis de parte del ARN ribosómico.

El ARN ribosómico (ARNr) es el más abundante de la célula. Está formado por una sola cadena de nucleótidos, aunque presenta zonas de doble hélice debido a su conformación tridimensional.

### **1.3. GNESES, CROMOSOMAS, CROMATINA Y NUCLEOSOMA**

En términos generales se habla de un conjunto de información que lleva consigo una instrucción en particular que generalmente codifica una proteína. El ADN que conforma los genes, almacena la información genética en el núcleo. Existen fragmentos de ADN que no codifican una proteína. Codifican otras moléculas como el ARN estructural y a esos también los debemos denominar genes. Además, se presentan situaciones en las que un fragmento de ADN es capaz de codificar más de una proteína. Entonces, ¿se debe hablar de un gen o de varios? En esas situaciones tendemos a confundirnos un poco con la terminología. Pero en general, se puede decir que un gen es un fragmento de ADN, que puede tener 100 pares de bases o hasta dos millones de pares de bases, como en el caso del gen de la distrofia muscular, el cual codifica una proteína en particular que tiene a su vez una función particular.

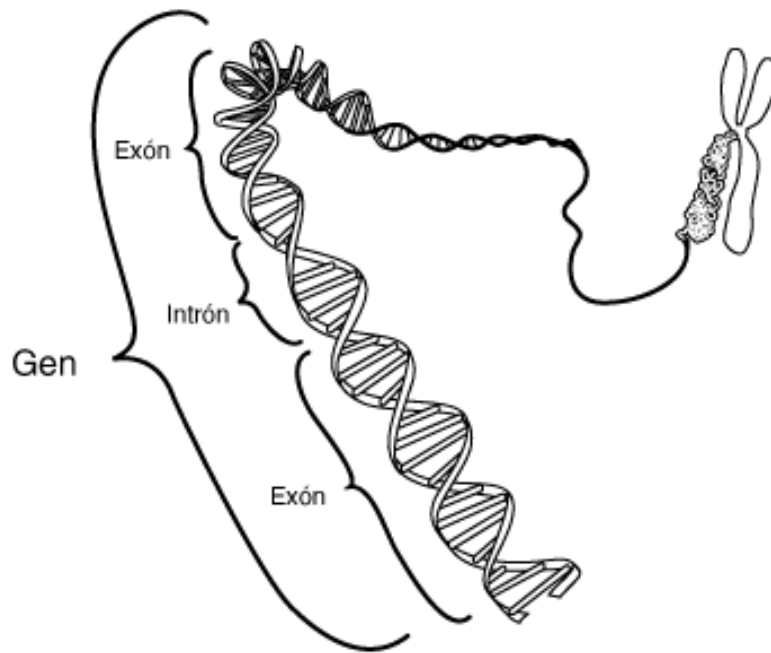


Imagen 13.

Fuente: <http://genome.gov/sglossary.cfm?ID=113&action=ver>

Todas las células vivas almacenan su información genética en estructuras llamadas cromosomas. Los cromosomas están constituidos de ADN y proteínas que están empacadas en forma compacta y al examinarlos microscópicamente parecen un hilo o una soga. En el caso de las células nucleadas, como las células humanas, los cromosomas están localizados dentro del núcleo. Diferentes organismos tienen distinto número de cromosomas. Los humanos tenemos 23 pares de cromosomas o sea 46 en total; 44 autosomas y 2 cromosomas sexuales. Cada progenitor aporta un cromosoma a cada par, así que los hijos adquieren la mitad de sus cromosomas de la madre y la mitad del padre. De esta manera se crea una nueva combinación genética, que representa la mezcla de ambos padres.

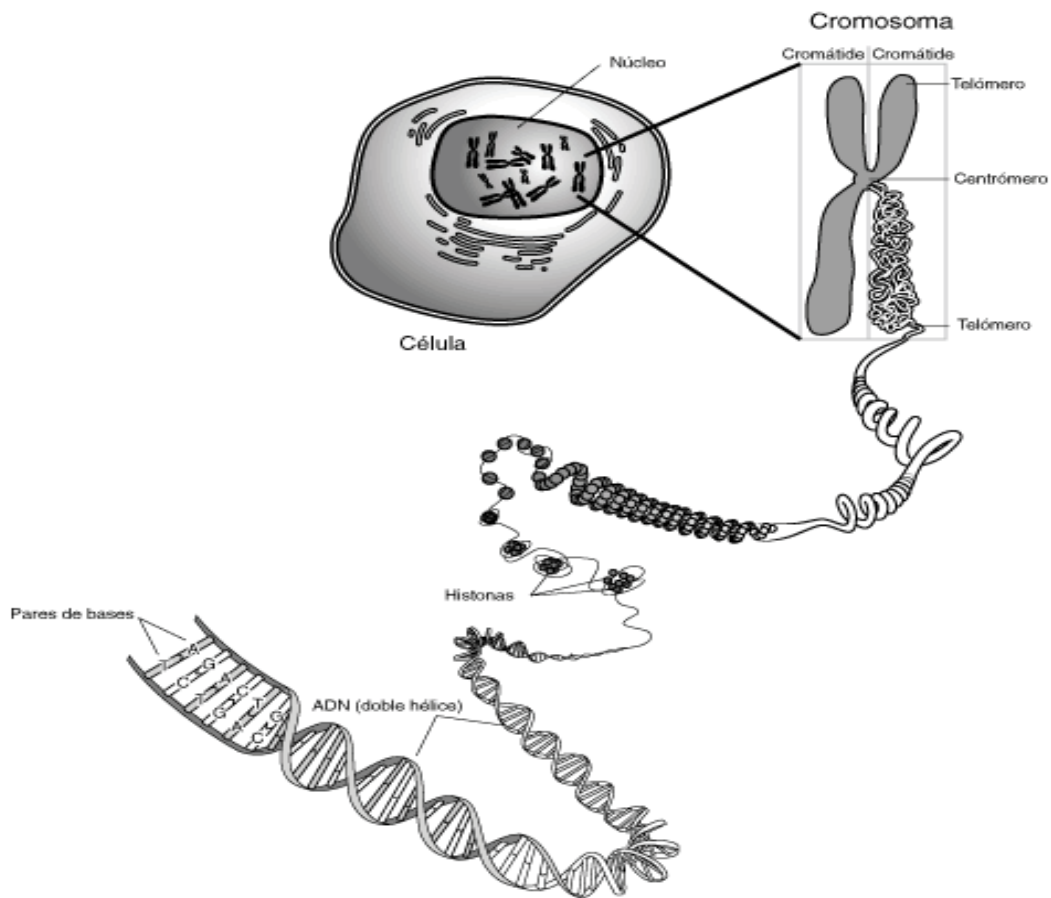


Imagen 14.

Fuente: <http://genome.gov/sglossary.cfm?ID=95&action=ver>

Cada molécula de ADN con sus proteínas asociadas recibe el nombre de cromosoma. La condensación del ADN en cromosomas sirve para varias funciones importantes:

Primero, el cromosoma es una forma compacta del ADN que cabe en la célula. Segundo, la condensación del ADN en cromosomas sirve para protegerlo de las lesiones. Las moléculas de ADN desnudas por completo son



relativamente inestables en las células. En cambio, el ADN cromosómica es extremadamente estable. Tercero, sólo el ADN condensado en un cromosoma puede ser transmitido con eficacia a ambas células hijas cada vez que una se divide. Por último, el cromosoma confiere una organización general a cada molécula de ADN. Esta organización facilita la expresión génica, al igual que la recombinación entre los cromosomas de los progenitores, que generan la diversidad observada entre los individuos diferentes de cualquier otro organismo. En las células eucariontes, una región dada de ADN.

## 1.4. REGULACIÓN DEL CICLO CELULAR<sup>72</sup>

La serie de sucesos por los cuales una célula se divide dando lugar a dos células hijas idénticas se denomina ciclo celular. Esta habilidad de reproducir réplicas exactas de sí misma es un componente esencial para la vida, no sólo de las células como ente individual, sino de los organismos que lo constituyen.

La proliferación celular tiene lugar de un modo controlado durante el desarrollo del cigoto, así como en el mantenimiento de las necesidades del organismo: por ejemplo, mientras las neuronas no se dividen, las células del hígado se dividen una vez al mes y los precursores de las células hematopoyéticas de la médula ósea se dividen de forma continuada, generando nuevas células que aseguran el mantenimiento del individuo. Parece claro, por tanto, que la maquinaria molecular que controla el ciclo y la división celular está finamente regulada y organizada, integrando simultáneamente múltiples señales, tanto extracelulares como intracelulares, con el objetivo de garantizar la fidelidad de esta difícil tarea.

La pérdida de la capacidad de respuesta a los mecanismos de control que regulan esta división es la característica que mejor define a la célula tumoral, ya que permite una proliferación más o menos descontrolada, que acaba matando al propio huésped.

---

<sup>72</sup> Este epígrafe ha sido elaborado según el manual *Biología Molecular de la Célula* 4ª ed (pp. 984-1026). Publicado por Ediciones Omega (2003).

Hay múltiples acciones moleculares que regulan el ciclo y sirven de freno, como las proteínas p15, p16 y p53, entre otras. La p53, detiene el ciclo celular enviando a la célula a reparar si el daño no es irreversible o a la apoptosis si no reúne las condiciones genéticas correctas para su replicación. Es la guardiana del ciclo celular.

Los cánceres surgen o por un exceso de la activación de los genes codificantes de proteínas estimuladoras (proto-oncogenes transformados en oncogenes) o por el fracaso de los genes encargados del freno, o alteración de los genes reparadores o de la apoptosis; o por combinación de e alguno de ellos, aunque al final se van acumulando mutaciones y puede haber fallo de todos.

El ciclo celular consta de cuatro etapas:  $G_1$ , S,  $G_2$ , M. en su conjunto, y aunque hay variaciones, el ciclo completo dura unas 24 horas.

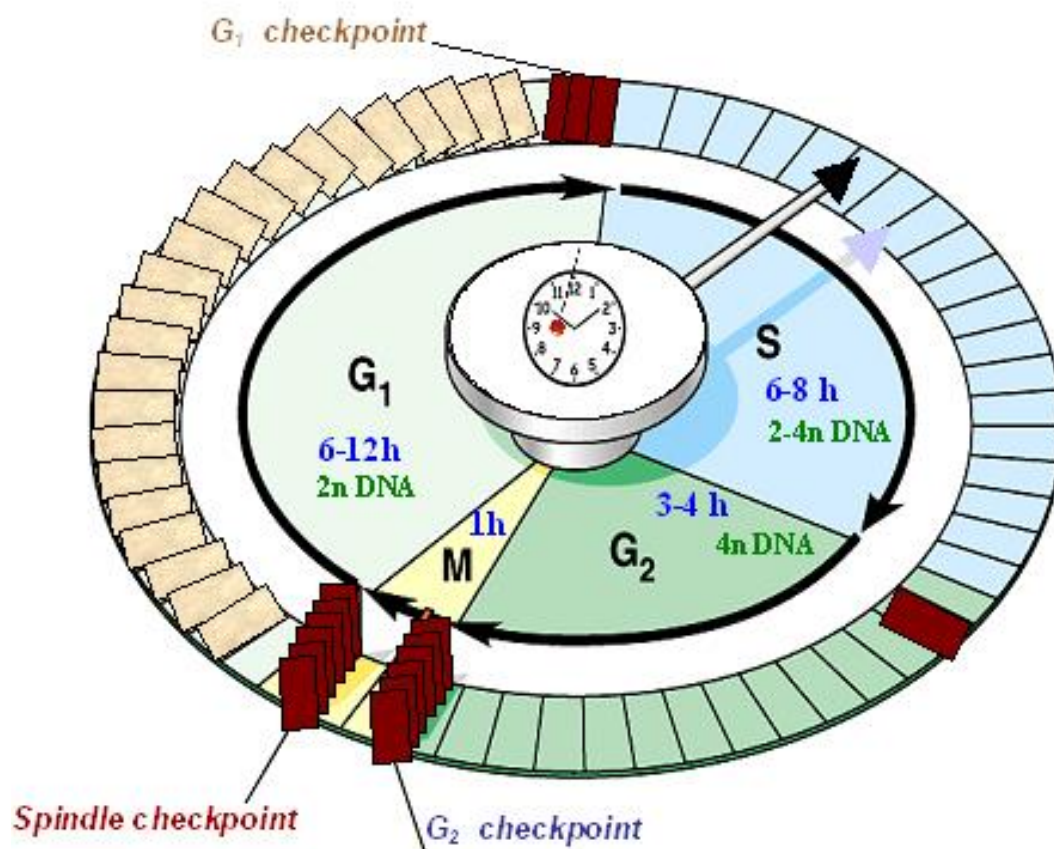


Imagen 15.

Fuente: <http://www.minet.uni-jena.de/~ibrahim/cell%20cycle%20pic.jpg>

La fase G<sub>1</sub> (del inglés gap o intervalo) es el período de 6 a 12 h que sigue a la división celular, previo a la síntesis o replicación del ADN durante este tiempo la célula dobla su tamaño y su masa debido a las síntesis continua de todos sus componentes como resultado de la transcripción y traducción de los genes que codifican las proteínas responsables de su fenotipo particular. Hay células que pueden parar su progresión hacia la división durante días, meses o años en estado de reposo sin aumento de masa, en lo que se denomina la fase G<sub>0</sub>. Las células en G<sub>0</sub> suelen actuar como una población en

reserva que en determinadas condiciones pueden reiniciar el ciclo celular y repoblar el tejido o el tumor. La continuación del ciclo o la entrada a  $G_0$ , depende de los estímulos derivados del microambiente (factores de crecimiento, factores nutritivos, etc.). En la fase  $G_1$  existe un punto de control llamado el punto de control de restricción R en el que la célula comprueba que ha generado la masa necesaria para seguir adelante y comienza la síntesis de ADN y, también, que las condiciones ambientales son favorables: presencia de nutrientes, de sales y temperatura adecuadas; y de factores que induzcan crecimiento adecuado sobre receptores en la membrana celular.

La fase S (de síntesis de ADN) corresponde al tiempo (6-8h) durante el cual se replica el ADN. Cada cromosoma pasa a tener dos cromátidas, es decir, dos moléculas de ADN de cadena doble, que son copia una de la otra.

El período comprendido entre la finalización de la replicación del ADN y el inicio de la división en la fase G2 (3-4 h). Durante ella, las células se preparan para la escisión de dos células hijas. En esta fase existe un punto de control G2-M, en la que la célula debe comprobar dos condiciones: una, que ha duplicado su masa de modo que puede dar lugar a dos células hijas, y otra, que ha completado la replicación del ADN, y sólo lo ha hecho una vez.

Finalmente, las células entran en la fase de mitosis M (1 h). Los cromosomas se condensan densamente. Los microtúbulos se organizan a partir de dos cuerpos polares que se sitúan a ambos extremos de la célula formando el huso acromático que va a servir como guía a los cromosomas.

Existe un punto de control M en la mitosis que sólo permite seguir adelante si todos los cromosomas están alineados sobre el huso.. si esto es así, una señal causa que las cromátidas hermanas se separen, continuando así el proceso mitótico.

Fases de la mitosis:

- Profase
- Metafase
- Anafase
- Telofase
- Citocinesis

## **ANEXO II. EL CÁNCER**

### **2.1. EL CANCER**

### **2.2. LA FORMACIÓN DEL TUMOR**

### **2.3. GENES IMPLICADOS EN EL CÁNCER**

#### **2.3.1. ONCOGENES Y PROTO-ONCOGENES**

#### **2.3.2. GENES SUPRESORES DE TUMORES**

##### **2.3.2.1. CLASIFICACIÓN FUNCIONAL DE LOS GENES SUPRESORES TUMORALES**

#### **2.3.3. LOS GENES DE REPARACIÓN DEL ADN**

### **2.4. MECANISMOS DE ACTIVACIÓN ONCOGÉNICA**

### **2.5. ANGIOGÉNES**

### **2.6. INVASIÓN Y METÁSTASIS**

### **2.7. LA APOTOSIS EN PROCESOS TUMORALES**

## 2.1. EL CÁNCER<sup>73</sup>

El cáncer a día de hoy es un tema de trascendencia social de gran impacto. Se calcula que en este siglo 1 de cada 3 españoles padecerán un cáncer, y aunque más del 50% podrá conseguir la curación (a la luz de los conocimientos actuales), casi el otro 50% morirá.

Los genes que controlan la división y proliferación celular son responsables de ello. El cáncer es un fenómeno altamente ligado al reloj biológico celular y por tanto al envejecimiento por lo que al aumentar la esperanza de vida de la población occidental aumentará a su vez la incidencia de procesos neoplásicos malignos. Estos genes tienen misión y características diferentes, de modo que mientras algunos tienen una acción de señalización positiva, otros son de naturaleza supresora y otros reparadores de genes de ADN. Cada vez los conocimientos al respecto son más sólidos, pero cada vez más complejos. Ya no se entiende tan sólo que el cáncer sea un profundo desajuste de la división celular, sino también de la muerte programada o apoptosis<sup>74</sup>.

Desgraciadamente el cáncer no es una enfermedad monogénica (interviene un sólo gen), de modo que el desideratum terapéutico de corregir el

---

<sup>73</sup> Este epígrafe ha sido elaborado según el manual *Cáncer: Genes y Nuevas Terapias* (pp. 06-22). Publicado por la Editorial Hélice (1997).

<sup>74</sup> Tipo de muerte celular en la que una serie de procesos moleculares en la célula conducen a su muerte. Este es el proceso normal mediante el cual el cuerpo se deshace de células innecesarias o anormales. El proceso de apoptosis puede estar impedido en las células cancerosas. También se llama muerte celular programada.



comportamiento de un solo gen es claramente insuficiente. Son muchos los genes involucrados. Por todo ello la terapia génica<sup>75</sup> se presenta, aunque de manera prometedora, como complemento necesario a otras terapéuticas al uso. Ello supone la entrada a escena de nuevas ramas de la medicina como la genómica<sup>76</sup> y de otras ciencias como la estadística, las matemáticas y la estadística.

El nombre de cáncer proviene del término griego Karkima dado por Hipócrates cuatro siglos antes de Cristo, y del equivalente latino cáncer, al crecimiento radial infiltrante de algunos tumores de mama a modo del cuerpo de un cangrejo.

Los científicos utilizan una variedad de nombres técnicos para distinguir los muchos diferentes tipos de carcinomas, sarcomas, linfomas y leucemias. En general, estos nombres son creados utilizando diferentes prefijos latinos que corresponden a la ubicación en donde el cáncer empezó su crecimiento descontrolado. Por ejemplo, el prefijo "osteo" significa hueso, por lo que un cáncer que se origina en el hueso se conoce como un osteosarcoma. Similarmente, el prefijo "adeno" significa glándula, por lo que un cáncer de células glandulares se conoce como adenocarcinoma--por ejemplo, un adenocarcinoma del seno. En la figura 1 se muestran los principales tipos prefijos a la hora de denominar el tipo de cáncer.

---

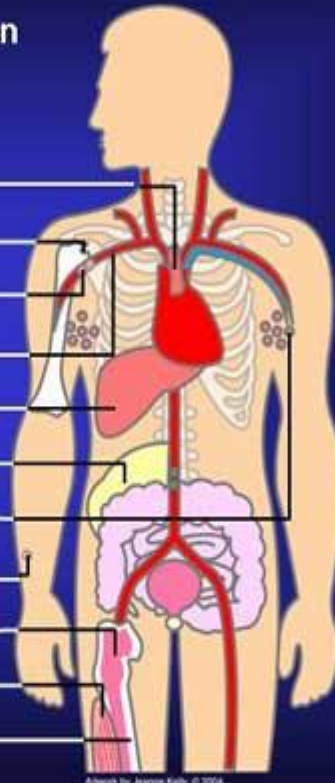
<sup>75</sup> Tratamiento que altera un gen. En los estudios de terapia génica para el cáncer, los investigadores están tratando de mejorar la capacidad natural del cuerpo para luchar contra la enfermedad o hacer que las células sean más sensibles a otras clases de terapia.

<sup>76</sup> Estudio del material genético completo, incluso los genes y sus funciones, en un organismo.

# Nombrando los Cánceres

Los Prefijos del Cáncer Señalan a la Ubicación

<b><i>Prefijo</i></b>	<b><i>Significado</i></b>
<b>adeno-</b>	glándula
<b>condro-</b>	cartílago
<b>eritro-</b>	glóbulos rojos
<b>hemangio-</b>	vasos sanguíneos
<b>hepato-</b>	hígado
<b>lipo-</b>	grasa
<b>linfo(á)-</b>	linfocito
<b>melano-</b>	célula de pigmento
<b>mielo-</b>	médula ósea
<b>mio-</b>	músculo
<b>osteo-</b>	hueso



NATIONAL  
CANCER  
INSTITUTE

Figura 1.

Fuente: Instituto Nacional del Cáncer

Aunque se reconocen más de un centenar de enfermedades neoplásicas y cada una representa sus peculiaridades específicas, es posible escribir pautas y patrones comunes en la génesis y desarrollo de la práctica de la totalidad de los tumores. En las última dos décadas se ha avanzado muchísimo en la elucidación de las bases moleculares del cáncer y se han hecho progresos significativos en la búsqueda de nuevas aproximaciones terapéuticas.

## 2.2. LA FORMACIÓN DEL TUMOR<sup>77</sup>

La supervivencia, proliferación y diferenciación de las células está regulada por señales externas. Las células proliferan dividiéndose (como se muestra en la figura 2) para formar células hijas con idéntico material genético en respuesta a agentes y que generalmente son proteínas<sup>78</sup> presentes en el suero que reciben el nombre de factores de crecimiento. La mayoría de los factores de crecimiento<sup>79</sup> activan la división celular, aunque otros son factores de inhibidores. En el organismo, las células normales proliferan sólo en respuesta a factores que estimulan su ciclo celular.

Por el contrario, las células cancerosas han perdido la restricción al crecimiento o, necesitando un menor número o concentraciones más reducidas de factores de crecimiento, o siendo totalmente independientes de ellos. Los factores de crecimiento actúan mediante su unión a moléculas específicas denominadas receptores que se encuentran en la membrana plasmática celular. Como consecuencia de la formación de estos complejos se inducen una serie de procesos bioquímicos conocidos genéricamente como transducción de la señal mitogénica<sup>80</sup>. Dicha señal llega al núcleo, donde finalmente regula la

---

<sup>77</sup> Este epígrafe ha sido elaborado según el manual *Biología Molecular de la Célula* 4ª ed (pp. 1313-1325). Publicado por Ediciones Omega (2003).

<sup>78</sup> Molécula orgánica rica en nitrógeno, compuesta esencialmente por aminoácidos. Todas las membranas celulares están formadas por proteínas, por lo que constituyen un compuesto fundamental.

<sup>79</sup> Sustancia producida por el cuerpo que regula la división celular y la supervivencia celular. Algunos factores de crecimiento se producen también en el laboratorio y se usan en la terapia biológica.

<sup>80</sup> La transducción de señal es el conjunto de procesos o etapas que ocurren de forma concatenada por el que una célula convierte una determinada señal o estímulo exterior, en otra señal o respuesta específica.

expresión génica y el ciclo de división celular. Si falla alguna de estas etapas, es decir; la unión del factor de crecimiento al receptor<sup>81</sup>, o la vía de señalización que va al núcleo de la célula se producen como consecuencia células anormales. Cuando una señal llega al núcleo de manera errónea produce mutaciones<sup>82</sup> en el material genético que serán transmitidas a las células hijas.

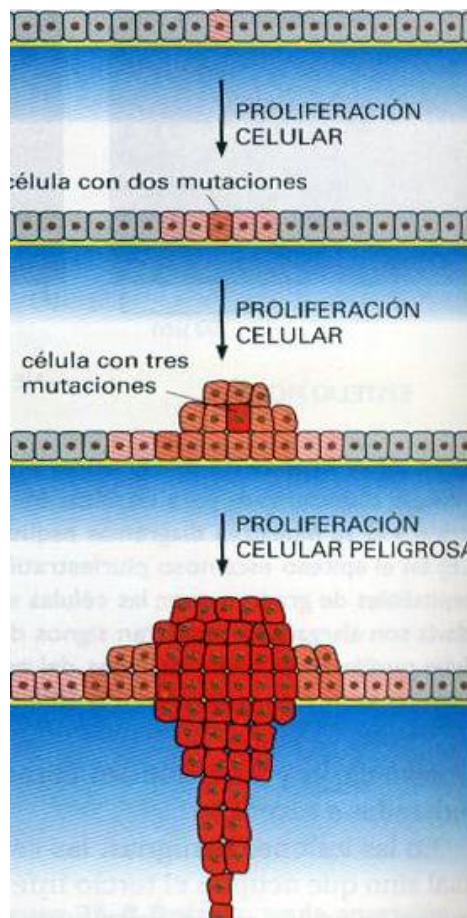


Imagen 2.

<sup>81</sup> Moléculas que se encuentran en la superficie de algunas células para unirse a los factores de crecimiento.

<sup>82</sup> Cualquier cambio en el ADN de una célula. Las mutaciones pueden ser causadas por errores durante la multiplicación de las células o por la exposición a sustancias del ambiente que dañan el ADN. Las mutaciones pueden ser nocivas, beneficiosas o no tener ningún efecto. Si se presentan en las células que producen los óvulos o los espermatozoides, pueden heredarse; si las mutaciones se presentan en otros tipos de células, no se heredan. Ciertas mutaciones pueden llevar a padecer de cáncer u otras enfermedades.

Fuente: Biología molecular de la célula. Cuarta edición, capítulo 23 pag. 1320

Los cambios genéticos involucrados en el cáncer resultan en proteínas alteradas que interrumpen la red de comunicación de la célula. En el cáncer, las proteínas alteradas a lo largo de muchos mecanismos diferentes causan que las señales sean confusas, interceptadas, amplificadas o dirigidas erradamente. Estos cambios secuestran lo que una vez era una comunicación normal y la utilizan para lograr alcanzar un crecimiento tumoral descontrolado.



Imagen 3.

Fuente: Instituto Nacional del Cáncer

El proceso de formación de un tumor a partir de una célula implica la acumulación sucesiva de alteraciones en los genes a lo largo de un período de

años de crecimiento hasta que se hace aparente. Durante este proceso dinámico de tipo evolutivo, aquellas células que van adquiriendo alteraciones en su material genético que les proporcionan una ventaja en cuanto al crecimiento, van siendo seleccionadas y serán mayoritarias en el tumor.

En la evolución del tumor pueden distinguirse tres etapas:

- Transformación neoplásica.
- Progresión tumoral in situ dependiente de angiogénesis<sup>83</sup>.
- Invasión y metástasis<sup>84</sup>.

---

<sup>83</sup> Formación de vasos sanguíneos. La angiogénesis tumoral es el crecimiento de vasos sanguíneos nuevos que los tumores necesitan para crecer. Esto ocurre porque el tumor descarga productos químicos.

<sup>84</sup> Diseminación del cáncer de una parte del cuerpo a otra. Un tumor formado por células que se han diseminado se llama "tumor metastásico" o "metástasis". El tumor metastásico contiene células que son como aquellas del tumor original (primario).

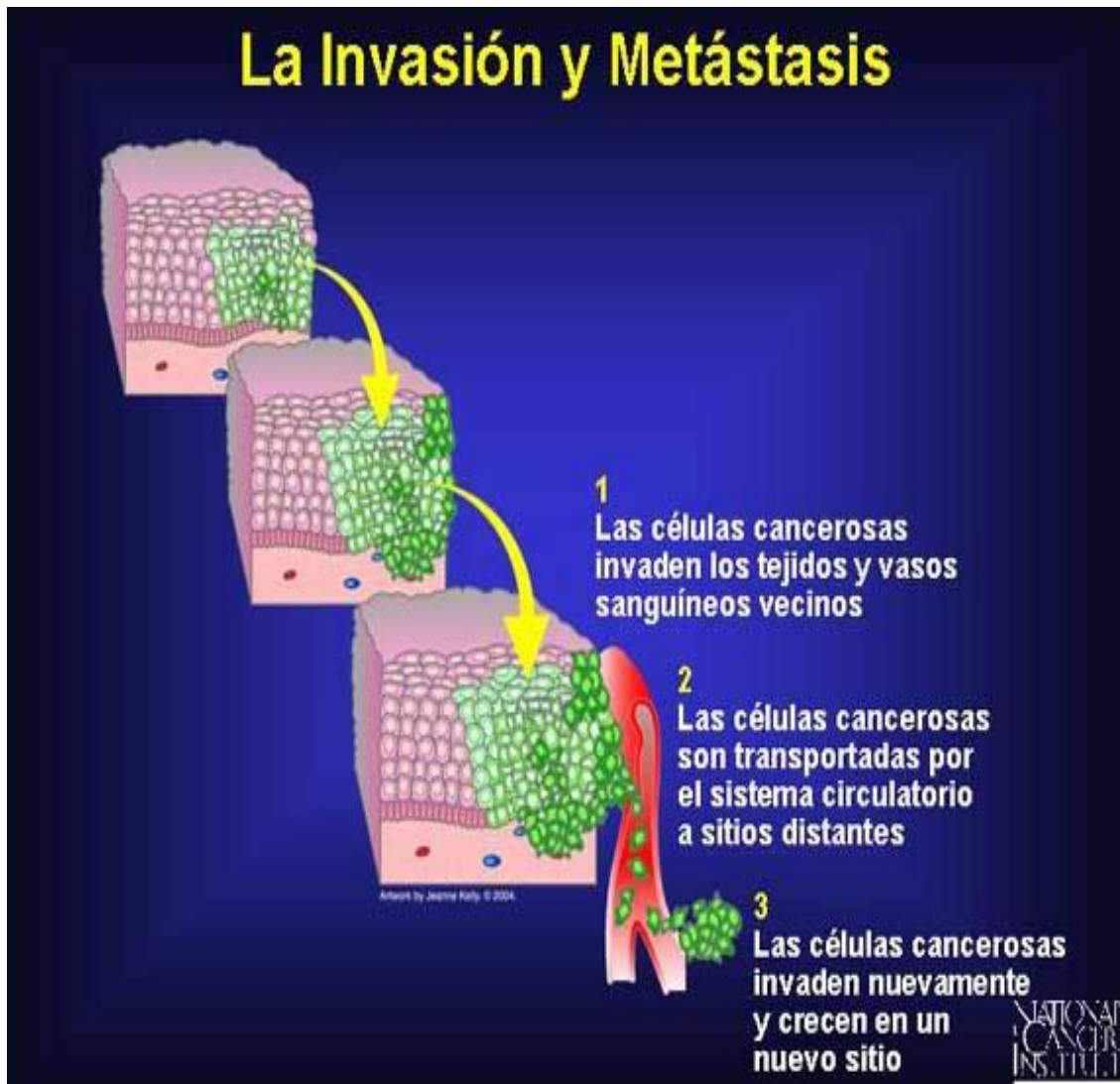


Imagen 4.

Fuente: Instituto Nacional del Cáncer

Cada una de estas tres etapas supone la acumulación de “fallos” en los diversos y sofisticados mecanismos de regulación del organismo. Estos “fallos” consisten básicamente en la acumulación de mutaciones en genes cuyos productos juegan funciones importantes en los mecanismos de regulación.

El cáncer aparece cuando las células de los tumores primarios que crecen en un determinado lugar, adquieren la capacidad de escaparse e invadir

otros tejidos, colonizarlos y crecer en ellos volviendo a generar tumores secundarios. Estos fenómenos de invasividad (capacidad para invadir) y metástasis implican nuevas mutaciones, ya que en condiciones normales las células ocupan un lugar determinado en el organismo. Así pues, son las mutaciones que controlan la proliferación celular y su localización en un sitio determinado las que causan la aparición de células cancerosas que pueden originar la destrucción de tejidos u órganos. Los cambios genéticos implicados en el desarrollo tumoral podemos resumirlos en:

- Capacidad de proliferar en ausencia de señales mitogénicas
- Capacidad para no responder a señales inhibitoras del crecimiento
- Capacidad de evitar la apoptosis
- Capacidad para reproducirse sin limitación
- Adquisición de la capacidad de generar angiogénesis
- Capacidad de invadir los tejidos colindantes y metastatizar.

Las células de nuestro organismo están dotadas de una maquinaria extraordinariamente precisa y sofisticada para controlar el ciclo celular. El ciclo celular consta, como se mencionó anteriormente de cuatro etapas: la fase de división activa o mitosis (M), la fase de replicación de la información genética (S) y dos fases “separadoras” (G1 y G2). Al período que comprenden las fases G1, S, G2 se le denomina interfase. Durante las fases G1 y G2 se han de producir las tomas de decisión que conduzcan a una nueva división celular o a al mantenimiento de la célula en estado quiescente (G0). Estas decisiones se



toman en función del resultado de las “inspecciones” que lleva a cabo la maquinaria de control del ciclo celular en los llamados puntos de control y chequeo para verificar que “todo marcha adecuadamente”

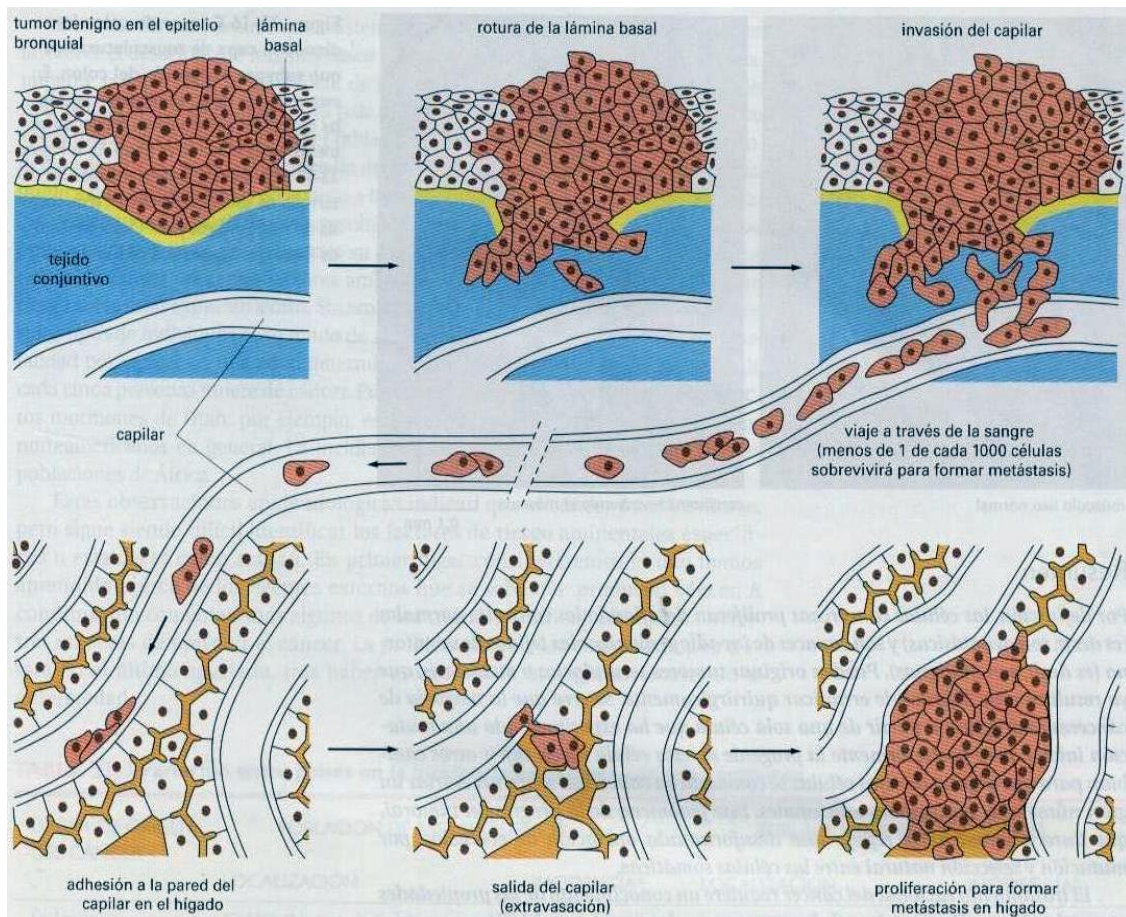


Imagen 5.

Fuente: Biología molecular de la célula. Cuarta edición, capítulo 23  
pag.1325

Por otra parte, nuestras células están también provistas de una maquinaria de control de la muerte celular programada genéticamente o

apoptosis. Se trata de un sofisticado sistema de salvaguardia del patrimonio genético por lo que las células dañadas son eliminadas para prevenir que el daño se extienda y, particularmente, para impedir que graves daños del ADN se perpetúen.

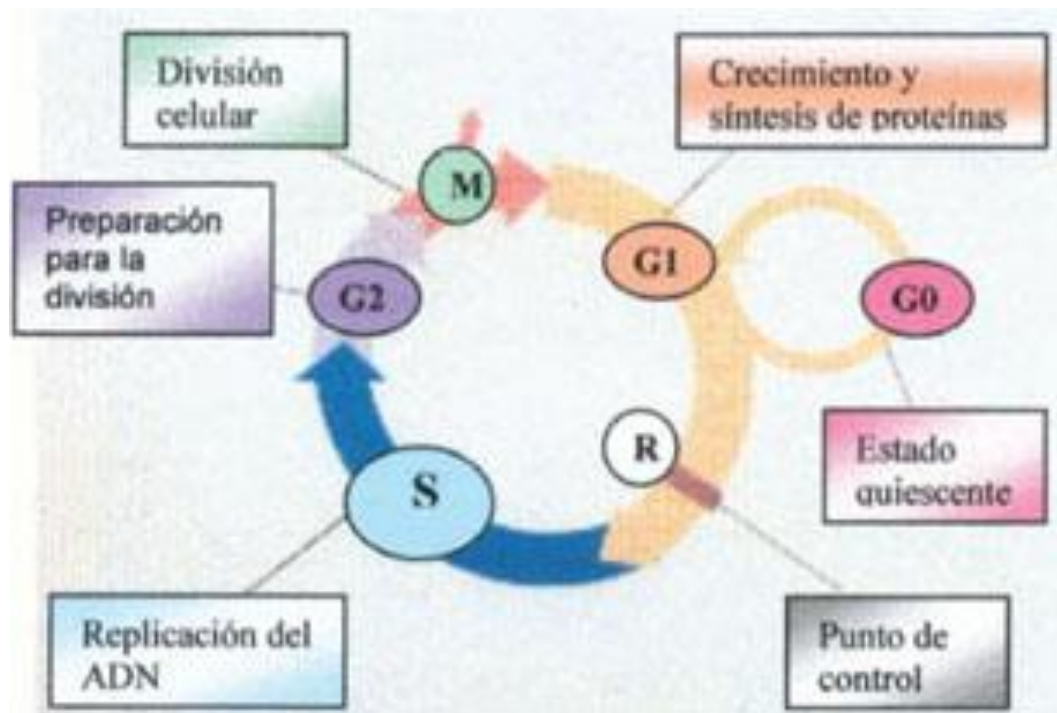


Imagen 6.

Fuente: <http://www.scielo.isciii.es/img/revistas/odonto/v21n6/287fig6.jpg>

## 2.3. GENES IMPLICADOS EN EL CÁNCER<sup>85</sup>

Las mutaciones causantes de los procesos que conducen a la aparición de un cáncer tienen lugar en tres tipos de genes: los proto-oncogenes<sup>86</sup>, los genes supresores de tumores y los genes de reparación del ADN.

### 2.3.1. Oncogenes y Proto-oncogenes

La nomenclatura utilizada en la literatura científica es a veces confusa. Como es lógico, nuestras células no tienen genes inductores de cáncer. Los oncogenes<sup>87</sup> son, en realidad, formas mutadas de genes normales, los proto-oncogenes. Es al mutar éstos, y originar proteínas con funciones alteradas que estimulan el crecimiento o la invasividad celular, cuando se convierten en oncogenes.

Los oncogenes están relacionados con los genes normales llamados proto-oncogenes, los cuales codifican los componentes del mecanismo de control de crecimiento normal de la célula. Algunos de estos componentes son factores de crecimiento, receptores, enzimas<sup>88</sup> señaladoras y factores de

---

<sup>85</sup> Este epígrafe ha sido elaborado según el manual *Biología Molecular de la Célula* 4ª ed (pp. 1333-1355). Publicado por Ediciones Omega (2003).

<sup>86</sup> Gen que participa en el crecimiento normal de las células. Las mutaciones (cambios) en un protooncogén pueden hacer que este se convierta en un oncogén, que puede hacer que se formen células cancerosas.

<sup>87</sup> Un oncogén es un gen anormal o activado que procede de la mutación o activación de un gen normal llamado protooncogén. Los oncogenes son los responsables de la transformación de una célula normal en una maligna que desarrollará un determinado tipo de cáncer.

<sup>88</sup> Las enzimas son proteínas complejas que producen un cambio químico específico en otras sustancias, sin que exista un cambio en ellas mismas.

transcripción<sup>89</sup>. Los factores de crecimiento se unen a receptores en la superficie de la célula, lo cual activa a las enzimas señaladoras dentro de la célula, las cuales a su vez activan a proteínas especiales conocidas como factores de transcripción dentro del núcleo de la célula. Los factores de transcripción activados "ponen en marcha" a los genes requeridos para el crecimiento y la proliferación celular.

La mayoría de las veces, incluso la mera expresión excesivamente elevada de la proteína normal codificada por un proto-oncogén es suficiente para inducir alteraciones en el crecimiento y morfología de la célula, proceso conocido como transformación celular. Se dice que los oncogenes son las formas activadas de los proto-oncogenes, consecuencia de mutaciones que causan una ganancia de función, es decir, un efecto biológico distinto del que tienen los proto-oncogenes normalmente. Por tanto se reserva el término de proto-oncogen al gen normal, y el de oncogenes a las formas mutadas de los mismos.

Los oncogenes resultan de la mutación de proto-oncogenes. Los oncogenes se asemejan a los proto-oncogenes en el sentido que codifican la producción de proteínas involucradas en el control del crecimiento. Sin embargo, los oncogenes codifican versiones alteradas (o cantidades excesivas) de estas proteínas de control del crecimiento, alterando de esta manera el mecanismo de señalamiento del crecimiento de las células.

---

<sup>89</sup> Los factores de transcripción son proteínas que coordinan y regulan la expresión de un gen o de un grupo de genes.

Al producir versiones o cantidades anormales de proteínas de control del crecimiento celular, los oncogenes hacen que el mecanismo de señalización del crecimiento de la célula se vuelva hiperactivo. Para explicar esto, usemos una simple metáfora, el mecanismo de control del crecimiento es como el pedal de gasolina (acelerador) de un automóvil. Entre más activo esté el mecanismo, más rápido se dividen y crecen las células. La presencia de un oncogen es como tener el acelerador pegado al piso del vehículo, causando que las células crezcan y se dividan continuamente. Una célula cancerosa puede contener uno o más oncogenes, lo cual significa que uno o más componentes en este mecanismo serán anormales.

#### 2.3.1.1. Clasificación funcional de los oncogenes

La mayoría de los oncogenes descritos pueden clasificarse en diversas familias funcionales, de acuerdo con la actividad bioquímica que presentan sus productos génicos. Todos los proto-oncogenes descritos hasta la fecha codifican proteínas implicadas en el complejísimo entramado de señalización que conduce, en cascadas sucesivas, al crecimiento, proliferación y diferenciación celular. Una posible clasificación, en función de las propiedades bioquímicas que presentan sus productos proteicos, es la siguiente:

- Factores de crecimiento
- Receptores
- Transductores de señales
- Activadores de transcripción

Factores de crecimiento: son polipéptidos (moléculas formadas por la unión de varios aminoácidos) secretados por las células, cuya función es unirse a un receptor y estimular el crecimiento celular. Cada factor de crecimiento se acopla a receptores específicos. Los factores de crecimiento funcionan como los primeros mensajeros que producen respuestas específicas en células diana de moléculas receptoras que los reconozcan. Una desregulación en los factores de crecimiento puede afectar tanto a la propia célula productora (estimulación autocrina), como a las células cercanas (estimulación paracrina) o distantes (estimulación endocrina). El primer gen identificado perteneciente a este grupo fue el PDGF (Platelet Derived Growth Factor).

Los receptores de los factores de crecimiento son proteínas de la membrana citoplasmática que presentan tres áreas o dominios de acción fundamental:

Dominio extracelular, es la zona de la proteína que se une al receptor

Dominio transmembrana, zona de la proteína localizada en el interior de la membrana citoplasmática.

Dominio intracelular, también conocida como zona tirosín kinasa, por ser la zona donde se activan las reacciones bioquímicas citoplasmáticas.

Por lo tanto los receptores de factores de crecimiento son moléculas que transmiten información unidireccionalmente desde la superficie de la célula, mediante la unión de un factor de crecimiento específico, hasta la zona citoplasmática, donde la cascada de reacciones bioquímicas iniciadas en la

zona tirosín kinasa. Termina en el núcleo celular estimulando la síntesis de material genético. Por lo tanto la alteración en estos factores podrá inducir a un crecimiento anormal.

Transductores de señales: la transducción de señales consiste en una cascada bioquímica que se inicia en el interior de la membrana citoplasmática y termina en el núcleo de la célula, estimulando la síntesis de ADN. Esta transmisión de señales está normalmente gobernada por procesos de fosforilación. Los oncogenes que actúan como transductores de señales pueden dividirse en diferentes grupos:

- Receptores de tirosín kinasa (abl, fes, fgr, lck, src, yes)
- Receptores de serina/treonina/kinasa (raf,mos,pim)
- Receptores de GTP (H-ras, K-ras, N-ras)
- Receptores G proteína (gsp,gip)

Los oncogenes más ampliamente estudiados en este grupo son los de la familia ras.

Activadores de la transcripción: los proto-oncogenes activadores de la transcripción, también conocidos como nucleares, se caracterizan por codificar proteínas que se unen al ADN activando la transcripción de otros genes, que estimulan la replicación del ADN de forma directa o indirecta. Dentro de este grupo se incluyen los genes supresores de tumores y genes reparadores del ADN.

El resto de proto-oncogenes tumorales más estudiados son: myc, myb, fos, rel, y ski, que mediante cambios en su regulación génica, son transformados en oncogenes y de esta forma contribuyen al crecimiento tumoral. El gen myc está muy relacionado con el crecimiento tumoral en distintas neoplasias. Referente a su actividad biológica, se ha demostrado que la proteína myc es capaz de abolir y suplir la acción de factores de crecimiento. Otra característica importante de esta proteína es su capacidad para mantener las células en estados diferenciados.

### 2.3.2. Genes supresores de tumores

Los genes supresores de tumores, cuya función normal es controlar el ciclo de división celular, evitando el crecimiento excesivo y promoviendo el mantenimiento de las características que especifican la localización de las células en un lugar determinado. Estos genes inducen la aparición de cáncer cuando, al mutar, dejan de expresarse o producen una proteína no funcional.

Los genes supresores de tumor son genes normales cuya ausencia puede conducir al cáncer. En otras palabras, si una célula pierde un par de genes supresores de tumor o si son inactivados por mutación, su falta de funcionamiento puede permitir que el cáncer se desarrolle. Los individuos que heredan un riesgo incrementado de desarrollar cáncer frecuentemente nacen con una copia defectuosa del gen supresor de tumor. Debido a que los genes ocurren naturalmente en pares (uno heredado de cada uno de los padres), un defecto heredado en una copia no causará el cáncer debido a que la otra copia normal aún funciona. Pero si la segunda copia se somete a la mutación, la



persona entonces puede desarrollar cáncer porque ya no existe alguna copia del gen que funcione.

En una visión quizá simplista pero no muy alejada de la realidad, los proto-oncogenes son los genes que dan lugar a aquellas proteínas componentes del ciclo de división que favorecen (empujan) el funcionamiento del ciclo y, por tanto, la multiplicación celular; mientras que los genes supresores originan proteínas que tienden a mantener bloqueado (frenan) dicho ciclo. La velocidad de crecimiento de una célula en cada momento es el resultado de la acción simultánea de los productores de ambos tipos de genes. Resulta claro que tanto una producción excesiva o anormal de las proteínas codificadas por los oncogenes, como una limitada, anormal o nula expresión de los productores de los genes supresores de tumores, pueden generar un crecimiento celular descontrolado.

Los genes supresores de tumor son una familia de genes normales que ordenan a las células producir proteínas que restringen el crecimiento y la división celulares. Ya que los genes supresores de tumor codifican las proteínas que disminuyen el crecimiento y la división celulares, la pérdida de estas proteínas permite que la célula crezca y se divida de forma incontrolada. Los genes supresores de tumor actúan como los frenos de un automóvil. La pérdida de la función del gen supresor de tumor es como tener frenos que no funcionan apropiadamente permitiéndole a la célula, por lo tanto, que se divida y crezca continuamente.

#### 2.3.2.1. Clasificación funcional de los genes supresores tumorales

Al igual que ocurre con los oncogenes, los genes supresores de tumor codifican proteínas cuyas funciones normales han de ser relevantes. De hecho, los productos de los genes supresores tumorales descritos hasta la fecha actúan modulando rutas de transmisión de señales que, a su vez, controlan el ciclo celular y/o la maquinaria genética de muerte celular programada.

Genes de proteínas con localización citosólica: como el gene APC que está asociado a la adenopoliposis de colon hereditaria, que puede conducir a carcinoma de colon con muy elevada incidencia.

Genes de proteínas nucleares: como el Rb y el famoso p53 ambos genes supresores de tumores.

Genes de proteínas con localización incierta o mixta: destacan el gen BRCA1, relacionado con la predisposición a desarrollar cáncer de mama y ovario; BRCA2, relacionado con cáncer de mama.

#### 2.3.3. Los genes de reparación del ADN

Son los responsables de subsanar los errores producidos en la incorporación de los nucleótidos durante la replicación del ADN, y de reparar las alteraciones inducidas en la molécula de ADN por radiaciones y agentes químicos. La mutación producida en los genes de reparación de errores durante la replicación del ADN conduce a la aparición de múltiples mutaciones en el genoma, lo que se conoce como el fenotipo mutador. Estas mutaciones tendrán efectos cancerígenos cuando impliquen a un proto-oncogen o genes supresores de tumores.

Se conocen unos 100 oncogenes, unos 30 genes supresores, y algunos genes implicados en la replicación del ADN, lo que representa una proporción muy pequeña de los aproximadamente 40.500 genes de nuestro genoma.

Es evidente el interés que tiene conocer cuales son y cómo funcionan los proto-oncogenes, los genes supresores y los genes de reparación del ADN, para desentrañar las bases de la transformación de las células normales en cancerosas, y poder así diseñar nuevas estrategias anticancerosas y mejorar las existentes.

## 2.4. MECANISMOS DE ACTIVACIÓN ONCOGÉNICA

La transformación de la célula de benigna a tumoral se debe a la existencia de cambios genéticos en la secuencia de nucleótidos de los genes que regulan el crecimiento de la diferenciación. De esta forma. Los proto-oncogenes son activados como genes tumorales u oncogenes. Existen tres mecanismos básicos por los que los proto-oncogenes son transformados en oncogenes:

- Mutaciones.
- Amplificaciones génicas.
- Translocaciones cromosómicas.

Mutaciones: la mutación supone un cambio hereditario en la secuencia de nucleótidos de un organismo y puede o no tener consecuencias en el fenotipo de la célula. El proceso por el cual se adquiere una mutación se conoce como mutagénesis. El tipo de mutación más frecuente es la mutación puntual, que supone el cambio de una base nitrogenada por otra, bien por la inserción o la delección de una base nitrogenada. Las mutaciones puntuales pueden ser de distintos tipos:

Transiciones: cuando se produce un cambio de una base púrica por otra púrica.

Transversiones: cambios de una base púrica por otra pirimidínica o viceversa.

Inserciones o deleciones: inclusión o pérdida de una base, con la consecuencia de pérdida de lectura del código genético.

Ampliación génica: la ampliación génica se refiere al incremento en el número de copias de un gen dentro del genoma. La ampliación conduce a un incremento del gen y generalmente se relaciona con un incremento de la proliferación celular.

Reordenamiento cromosómico: este tipo de mecanismo de activación oncogénica, se produce frecuentemente en los cánceres hematológicos<sup>90</sup> y con menos incidencia en los cánceres sólidos. Los reordenamientos cromosómicos consisten más en traslocaciones que en inversiones cromosómicas. La transformación celular benigna en maligna se produce por dos mecanismos distintos:

- Activación transcripcional del proto-oncogen
- Fusión génica

---

<sup>90</sup> Algunos tipos de cáncer en la sangre.

## **2.5. ANGIOGÉNES**

La angiogénesis tumoral es la proliferación de una red de vasos sanguíneos que penetra dentro de crecimientos cancerosos, proporcionando nutrientes y oxígeno y removiendo productos de desecho. La angiogénesis tumoral empieza en realidad cuando las células cancerosas del tumor liberan moléculas que envían señales al tejido huésped normal vecino. Esta señalización activa a ciertos genes en el tejido huésped que, a su vez, producen proteínas para estimular el crecimiento de nuevos vasos sanguíneos.

Durante las últimas décadas se ha sugerido que la angiogénesis, o desarrollo de nuevos vasos sanguíneos, era una importante diana para la terapia antineoplásica. Este potencial no se había desarrollado hasta hace poco tiempo, cuando datos clínicos demostraron que un anticuerpo monoclonal humanizado, el bevacizumab, que actúa sobre la molécula proangiogénica crítica, el factor de crecimiento del endotelio vascular (VEGF), puede prolongar la supervivencia de los pacientes con cáncer colorrectal metastásico cuando se emplea como fármaco de primera línea combinado con quimioterapia.

## 2.6. INVASIÓN Y METÁSTASIS<sup>91</sup>

Un tumor localizado puede ser eliminado con relativa facilidad mediante cirugía en la mayoría de los casos. Sin embargo, la mayoría de las células que componen los tumores malignos presentan la capacidad de diseminarse y dan lugar a múltiples focos tumorales (metástasis) en otros lugares del organismo. Este proceso imposibilita una solución quirúrgica del problema y a pesar de los avances logrados en la prevención precoz y en la terapia antitumoral, constituye todavía la causa de muerte de un porcentaje muy elevados de pacientes con cáncer.

La fase inicial del proceso tumoral, la transformación neoplásica, conduce a la formación del tumor primario y es, sin duda, la etapa de la progresión tumoral a cuyo estudio se han dedicado más esfuerzos. Gracias a dichos esfuerzos se ha logrado acumular una gran cantidad de información molecular de los procesos que conducen a la transformación de una célula en tumoral.

El desarrollo de metástasis es el resultado de una cascada de acontecimientos interrelacionados que implican constantes interrelaciones entre las células tumorales y el organismo hospedador.

---

<sup>91</sup> Este epígrafe ha sido elaborado según el manual *Cáncer: Genes y Nuevas Terapias* (pp. 260-280). Publicado por la Editorial Hélice (1997).

El tumor primario, aunque integrado por células descendiente de una única célula transformada, adquiere rápidamente un variable de heterogeneidad como resultado de la acumulación progresiva de mutaciones. Como consecuencia, en este tumor primario existirán diferentes clones con capacidad variable de superar los distintos pasos de la cascada metastásica; de tal manera que existe un estricto mecanismo de selección natural que favorece a aquellas células con características más agresivas. Algunas de estas células serán capaces de inducir la formación de nuevos vasos sanguíneos (angiogénesis), abandonar el tumor primario y atravesar la matriz extracelular que les rodea (invasión), alcanzar el torrente circulatorio, sobrevivir en el mismo y abandonarlo atravesando las paredes de los capilares localizados en lugares distantes del organismo (diseminación), invadiendo nuevamente la matriz extracelular y estableciendo en la nueva localización un tumor secundario (colonización). Finalmente, este tumor secundario, originalmente microscópico requerirá la formación de nuevos vasos sanguíneos que le permitan crecer y diseminar sus células para dar lugar a la aparición de nuevas metástasis. Cada uno de los pasos descritos conlleva una serie de alteraciones moleculares y genéticas en las células tumorales que podrían constituir nuevas dianas terapéuticas.

- Podemos decir que fases de crecimiento de un tumor son las siguientes:
- Invasión
- Diseminación
- Escape de las defensas del huésped



**1. Invasión.** Se ha propuesto una hipótesis trifásica para describir los fenómenos que tienen lugar durante la invasión de la matriz extracelular.

El primer paso sería la adhesión celular. En condiciones normales es un freno contra la migración de las células. Hay dos tipos de adhesión: adhesión intracelular (célula-célula) y la adhesión a la matriz extracelular que permite a las células normales sobrevivir y proliferar. Desde hace años se sabe que, en cultivo, las células no pueden multiplicarse hasta que se une a una superficie, fenómeno que se conoce como dependencia de anclaje. Muchos tipos de células, si se les niega el punto de anclaje, no se limitan a suspender la proliferación, sino que se suicidan. Las células, para sobrevivir, necesitan anclarse a un sitio determinado. Pues bien, las células tumorales son capaces de vivir sin la adhesión a sus células vecinas y adherirse a una matriz que no les corresponde o sobrevivir sin unirse a ninguna matriz.

En una segunda fase, las células tumorales secretarían enzimas hidrolíticas<sup>92</sup>, o inducirían su secreción en las células del huésped<sup>93</sup>, que pueden degradar la matriz. Las lisis inducidas por estas enzimas ocurrirían en una zona muy localizada y cercana a la superficie tumoral.

El tumor es capaz, cuando está muy diferenciado de fabricar su propia

---

<sup>92</sup> Las enzimas hidrolíticas aceleran las reacciones en las que una sustancia se rompe en componentes más simples por reacción con moléculas de agua.

<sup>93</sup> Células invadidas por el tumor.

matriz, basado en un crecimiento anómalo del tejido conectivo<sup>94</sup>, fenómeno llamado desmoplasia<sup>95</sup>, que permite un mejor tránsito de las células tumorales.

El tercer paso sería la locomoción de la célula tumoral hacia la zona de la matriz modificada por la proteólisis<sup>96</sup>. Diferentes agentes favorecen el movimiento de las células en diversas direcciones: hacia estructuras vecinas, linfáticas<sup>97</sup> o vasos capilares<sup>98</sup>.

**2. Diseminación.** Partiendo de los conceptos arriba expuestos, habría 3 tipos de diseminación:

- Directa
- Vía linfática
- Vía hematológica

Directa: cuando los tumores crecen y penetran en las cavidades corporales pueden liberar células o grupos celulares, que viajan por estas para terminar formando siembras en las superficies serosas de otros órganos. Los tumores pulmonares que penetran en la pleura pueden dar lugar a metástasis parietales o viscerales. Por otra parte

---

<sup>94</sup> El tejido conectivo participa de la cohesión o separación de los diferentes elementos tisulares que componen los órganos y sistemas.

<sup>95</sup> Fenómeno de producción excesiva de tejido conectivo, pobre en células y rico en fibras de colágeno, que suele acompañar a algunos tumores malignos epiteliales.

<sup>96</sup> La proteólisis es la degradación de proteínas ya sea mediante enzimas específicas, llamadas proteasas, o por medio de digestión intramolecular.

<sup>97</sup> Del sistema linfático.

<sup>98</sup> Los capilares sanguíneos son los vasos sanguíneos de menor diámetro, están formados sólo por una capa de tejido, lo que permite el intercambio de sustancias entre la sangre y las sustancias que se encuentran alrededor de ella.

pueden invadir territorios adyacentes por invasión directa como es el caso del cáncer de estómago al páncreas o al hígado.

Vía linfática: el proceso de infiltración y de expansión de los tejidos del huésped ocurre por la penetración en los pequeños vasos linfáticos. Se forma émbolos que entran en la circulación y pueden quedar detenidos en el primer nódulo linfático o bien superarlos. El problema fundamental es si los ganglios linfáticos reacciona inmunológicamente frenando el avance del tumor o únicamente actúan como meros canales de tránsito.

Vía hematógena: las células malignas frecuentemente atraviesan las paredes de los capilares, pero rara vez pueden atravesar las arterias o arteriolas, ricas en elastina. Esto no depende de la resistencia mecánica a la invasión, si no que la elastina posee inhibidores de las proteasas<sup>99</sup> tumorales. Los tumores malignos no producen sus propios vasos sanguíneos, pero inducen el crecimiento de nuevos capilares. Una vez penetran en el torrente, pueden ser transportados por mecanismos pasivos o adherirse allí, crecer y después soltar émbolos. La presencia de células tumorales en el torrente sanguíneo no significa la presencia de metástasis, la mayoría son eliminadas.

**3. Escape de las defensas del huésped.** Las células tumorales deben ser capaces de superar las defensas del huésped en cada una de las

---

<sup>99</sup> Substancia en la sangre que destruye las proteínas.

fases de la cascada metastásica<sup>100</sup>. Se sabe que en la superficie de las células transformadas se expresan determinados antígenos<sup>101</sup> nuevos que deberían permitir al sistema inmune reconocer a esas células y producir una respuesta frente a ellas. Sin embargo, la neoplasia puede escapar a este sistema de control mediante varios mecanismos, de entre los cuales destacan:

- Generación de células T supresoras<sup>102</sup>. El crecimiento tumoral va acompañado de la incapacidad de dichas células para segregar linfocitos<sup>103</sup> que activen otras células de defensa.
- Liberación de sustancias inmunosupresoras<sup>104</sup> por el tumor. En estadios más avanzados, los tumores pueden producir sustancias con capacidad de inhibir respuestas inmunitarias de tipo hormonal y celular.
- Presentación inadecuada de antígenos.

---

<sup>100</sup> Conjunto procesos bioquímicos necesarios para que se produzca la metástasis.

<sup>101</sup> Así se llama a toda sustancia extraña al organismo que, una vez introducida en el cuerpo, es capaz de provocar anticuerpos.

<sup>102</sup> Tipo de célula inmunitaria que impide la acción de algunos otros tipos de linfocitos para que el sistema inmunitario no se vuelva sobreactivo. Las células T supresoras están en estudio para el tratamiento de cáncer. La célula T supresora es un tipo de glóbulo blanco y un tipo de linfocito.

<sup>103</sup> Tipo de célula inmunitaria elaborada en la médula espinal; se encuentra en la sangre y el tejido linfático. Los dos tipos de linfocitos son los linfocitos B y los linfocitos T. Los linfocitos B elaboran anticuerpos y los linfocitos T ayudan a destruir las células tumorales y ayudan a controlar las respuestas inmunitarias. Un linfocito es un tipo de glóbulo blanco.

<sup>104</sup> sustancias o fármacos capaces de inhibir o reducir los mecanismos inmunitarios.

## 2.7. LA APOTOSIS EN PROCESOS TUMORALES

El mantenimiento de los tejidos celulares en organismos multicelulares es un equilibrio dinámico entre procesos de proliferación/diferenciación celular y eliminación de células por apoptosis. La apoptosis, denominada frecuentemente como muerte celular programada, es un proceso caracterizado por una serie bastante constante de cambios morfológicos en la célula, como es la disminución del volumen del núcleo, la fragmentación y condensación de la cromatina y finalmente la desaparición de la membrana celular.

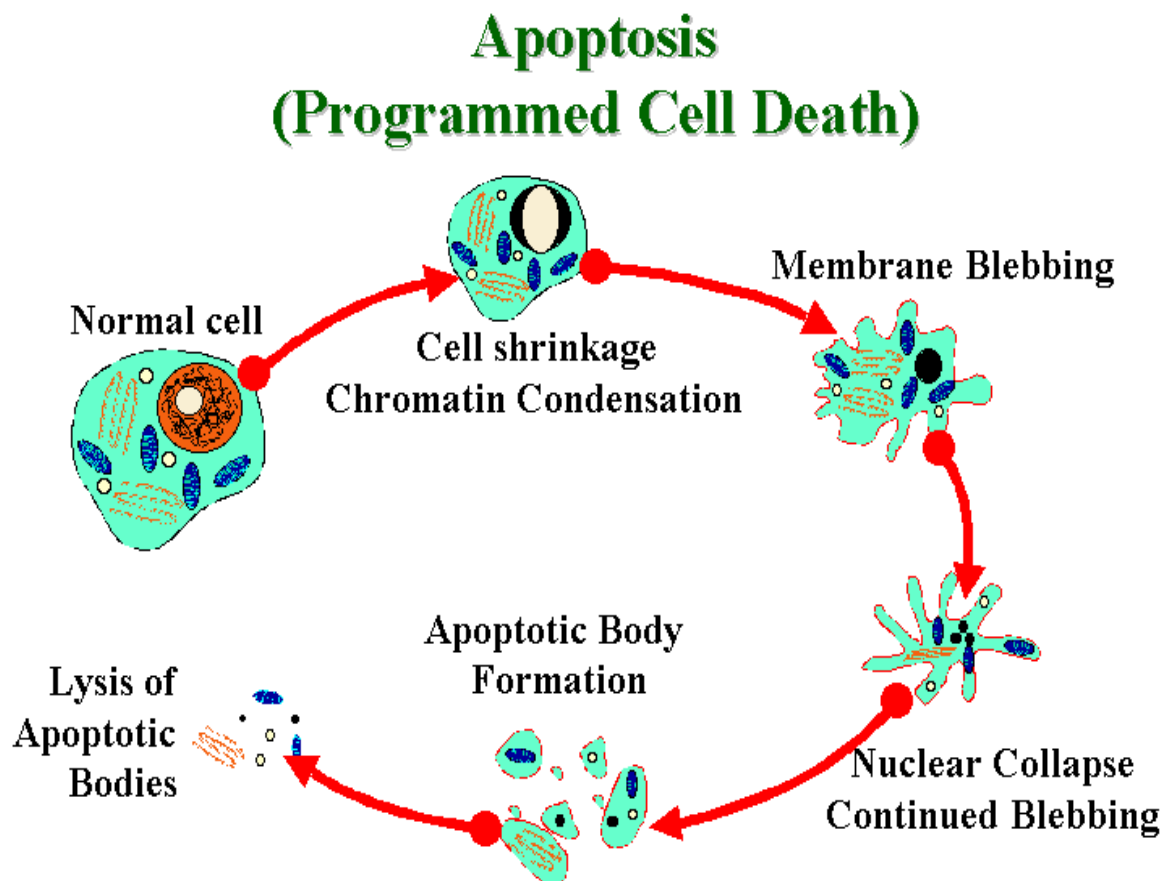


Imagen 13.

Fuente: <http://wwwmicro.msb.le.ac.uk/3035/kalmakoff/baculo/pics/Apoptosis.gif>

Para que el proceso de muerte celular se ponga en marcha es necesario determinadas señales, tanto extra como intracelulares. Esta forma de muerte celular desempeña un papel crucial no sólo durante el mantenimiento de los tejidos en tejidos adultos sino también en el desarrollo embrionario y es posiblemente un componente importante de alteraciones patológicas como el cáncer.

## **ANEXO III. RELACIÓN DEL CONJUNTO SELECCIONADOS**

3.1. GENES SELECCIONADOS COMUNES EN AMBOS ESTUDIOS

3.2. GENES SELECCIONADOS EN ESTE TRABAJO NO COMUNES AL  
ESTUDIO DEL GRUPO DE LA DRA. PÉREZ-VILLAMIL

### 3.1. GENES SELECCIONADOS COMUNES EN AMBOS ESTUDIOS

La siguiente tabla muestra en la primera columna el identificador del gen empleado en el análisis. En la segunda columna el símbolo del gen que corresponde con el identificador y, por último, en la tercera columna una breve descripción del gen.

De los 306 genes que comparten ambos estudios, 70 no están aún clasificados por las bases de datos de genes. Estos genes no tienen información ni en la columna *Símbolo* ni en la columna *Descripción*.

Identif icador	Sí mbolo	Descripción del Gen
NM_19 4272	R BPMS2	RNA binding protein with multiple splicing 2; RNA-binding protein with multiple splicing 2; Homo sapiens RNA binding protein with multiple splicing 2 (RBPMS2), mRNA.
NM_17 8493	N OTUM	notum pectinacylesterase homolog (Drosophila); Homo sapiens notum pectinacylesterase homolog (Drosophila) (NOTUM), mRNA.; Notum pectinacylesterase homolog (Drosophila)
NM_00 0064	C 3	complement component 3; synonyms: ASP, CPAMD1; acylation-stimulating protein cleavage product; complement component C3; Homo sapiens complement component 3 (C3), mRNA.; Complement component 3
NM_15 2358	C 19orf41	chromosome 19 open reading frame 41; synonyms: SCRL, MGC33947, PLAL6978, PRO21961; Homo sapiens chromosome 19 open reading frame 41 (C19orf41), mRNA.; Chromosome



		19 open reading frame 41
NM_002842	P TPRH	protein tyrosine phosphatase, receptor type, H; synonyms: SAP-1, FLJ39938, MGC133058, MGC133059; Homo sapiens protein tyrosine phosphatase, receptor type, H (PTPRH), mRNA.; Protein tyrosine phosphatase, receptor type, H
NM_003500	A COX2	acyl-Coenzyme A oxidase 2, branched chain; synonyms: BCOX, BRCOX, THCCox, BRCACOX; Peroxisomal branched chain acyl-CoA oxidase; Trihydroxycoprostanoyl-CoA oxidase; THCA-CoA oxidase; Homo sapiens acyl-Coenzyme A oxidase 2, branched chain (ACOX2), mRNA.; Acyl-Coenzyme A oxidase 2, branched chain
NM_003881	W ISP2	WNT1 inducible signaling pathway protein 2; synonyms: CCN5, CT58, CTGF-L; wnt-1 signaling pathway protein 2; connective tissue growth factor-like protein; Homo sapiens WNT1 inducible signaling pathway protein 2 (WISP2), mRNA.
NM_002010	F GF9	fibroblast growth factor 9 (glia-activating factor); synonyms: GAF, HBFG-9, MGC119914, MGC119915; Homo sapiens fibroblast growth factor 9 (glia-activating factor) (FGF9), mRNA.; Fibroblast growth factor 9 (glia-activating factor)
NM_021615	C HST6	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6; synonym: MCDL1; macular dystrophy, corneal, 1; carbohydrate sulfotransferase 6; corneal N-acetylglucosamine 6-sulfotransferase; Homo sapiens carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6 (CHST6), mRNA.; Carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6
NM_02	S	sal-like 4 (Drosophila); synonyms: DRRS,

0436	ALL4	HSAL4, ZNF797, MGC133050, dJ1112F19.1; sal (Drosophila)-like 4; Homo sapiens sal-like 4 (Drosophila) (SALL4), mRNA.; Sal-like 4 (Drosophila)
NM_004417	D USP1	dual specificity phosphatase 1; synonyms: HVH1, MKP1, CL100, MKP-1, PTPN10; serine/threonine specific protein phosphatase; Homo sapiens dual specificity phosphatase 1 (DUSP1), mRNA.; Dual specificity phosphatase 1
NM_138290		Rap2-binding protein 9; synonyms: RPIP9, FLJ30671, MGC26655; Rap2-interacting protein 9; Rap2 binding protein 9; Homo sapiens Rap2-binding protein 9 (RPIB9), mRNA.
NM_138455	C THRC1	collagen triple helix repeat containing 1; Homo sapiens collagen triple helix repeat containing 1 (CTHRC1), mRNA.; Collagen triple helix repeat containing 1
NM_002164	IN DO	indoleamine-pyrrole 2,3 dioxygenase; synonyms: IDO, CD107B; Indoleamine 2,3-dioxygenase; indole 2,3-dioxygenase; Homo sapiens indoleamine-pyrrole 2,3 dioxygenase (INDO), mRNA.; Indoleamine-pyrrole 2,3 dioxygenase
NM_021010	D EFA5	defensin, alpha 5, Paneth cell-specific; synonyms: DEF5, HD-5, MGC129728; defensin 5; defensin, alpha 5, preproprotein; Homo sapiens defensin, alpha 5, Paneth cell-specific (DEFA5), mRNA.; Defensin, alpha 5, Paneth cell-specific
NM_174921	C 4orf34	chromosome 4 open reading frame 34; synonym: FLJ13289; Homo sapiens chromosome 4 open reading frame 34 (C4orf34), mRNA.; Chromosome 4 open reading frame 34
NM_004791	IT GBL1	integrin, beta-like 1 (with EGF-like repeat domains); synonyms: OSCP, TIED; ten integrin EGF-

		like repeat domains protein; ITGBL1, integrin beta-like 1; Homo sapiens integrin, beta-like 1 (with EGF-like repeat domains) (ITGBL1), mRNA.; Integrin, beta-like 1 (with EGF-like repeat domains)
NM_007069	H RASLS3	HRAS-like suppressor 3; synonyms: HREV107, HREV107-3, MGC118754, H-REV107-1; Homo sapiens HRAS-like suppressor 3 (HRASLS3), mRNA.
NM_002286	L AG3	lymphocyte-activation gene 3; synonym: CD223; Homo sapiens lymphocyte-activation gene 3 (LAG3), mRNA.; Lymphocyte-activation gene 3
NM_001928	C FD	complement factor D (adipsin); synonyms: DF, ADN, PFD; properdin factor D; C3 convertase activator; D component of complement (adipsin); Homo sapiens complement factor D (adipsin) (CFD), mRNA.; Complement factor D (adipsin)
NM_005410	S EPP1	selenoprotein P, plasma, 1; synonyms: SeP, SELP; isoform 1 precursor is encoded by transcript variant 1; Homo sapiens selenoprotein P, plasma, 1 (SEPP1), transcript variant 1, mRNA.; Selenoprotein P, plasma, 1
NM_002192	IN HBA	inhibin, beta A (activin A, activin AB alpha polypeptide); synonyms: EDF, FRP; Inhibin, beta-1; Homo sapiens inhibin, beta A (activin A, activin AB alpha polypeptide) (INHBA), mRNA.; Inhibin, beta A (activin A, activin AB alpha polypeptide)
NM_000111	S LC26A3	solute carrier family 26, member 3; synonyms: CLD, DRA; down-regulated in adenoma; congenital chloride diarrhea; Homo sapiens solute carrier family 26, member 3 (SLC26A3), mRNA.; Solute carrier family 26, member 3
NM_00	C	chemokine (C-C motif) ligand 19; synonyms:

6274	CL19	ELC, CKb11, MIP3B, MIP-3b, SCYA19, MGC34433; small inducible cytokine subfamily A (Cys-Cys), member 19; CC chemokine ligand 19; macrophage inflammatory protein 3-beta; beta chemokine exodus-3; CK beta-11; EBI1-ligand chemokine; Homo sapiens chemokine (C-C motif) ligand 19 (CCL19), mRNA.; Chemokine (C-C motif) ligand 19
AF067 420	IG HA1	Homo sapiens SNC73 protein (SNC73) mRNA, complete cds.; immunoglobulin heavy constant alpha 1; Immunoglobulin heavy constant alpha 1
NM_00 5409	C XCL11	chemokine (C-X-C motif) ligand 11; synonyms: IP9, H174, IP-9, b-R1, I-TAC, SCYB11, SCYB9B, MGC102770; small inducible cytokine subfamily B (Cys-X-Cys), member 9B; small inducible cytokine subfamily B (Cys-X-Cys), member 11; small inducible cytokine B11; Homo sapiens chemokine (C-X-C motif) ligand 11 (CXCL11), mRNA.; Chemokine (C-X-C motif) ligand 11
NM_02 4087	A SB9	ankyrin repeat and SOCS box-containing 9; synonyms: MGC4954, FLJ20636, DKFZp564L0862; isoform 2 is encoded by transcript variant 2; Homo sapiens ankyrin repeat and SOCS box-containing 9 (ASB9), transcript variant 2, mRNA.; Ankyrin repeat and SOCS box-containing 9
NM_00 3326	T NFSF4	tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated glycoprotein 1, 34kDa); synonyms: GP34, CD252, OX4OL, TXGP1, CD134L, OX-40L; tax-transcriptionally activated glycoprotein 1 (34kD); glycoprotein 34 kd; OX40 ligand; OX40 antigen ligand; CD134 ligand; Homo sapiens tumor necrosis factor (ligand)

		superfamily, member 4 (tax-transcriptionally activated glycoprotein 1, 34kDa) (TNFSF4), mRNA.; Tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated glycoprotein 1, 34kDa)
NM_02 1800	D NAJC12	DnaJ (Hsp40) homolog, subfamily C, member 12; synonyms: JDP1, RP11-57G10.2; isoform a is encoded by transcript variant 1; J domain containing protein 1; J domain protein 1; J domain containing protein 1 (JDP1); Homo sapiens DnaJ (Hsp40) homolog, subfamily C, member 12 (DNAJC12), transcript variant 1, mRNA.
NM_01 7716	M S4A12	membrane-spanning 4-domains, subfamily A, member 12; synonym: Ms4a10; Homo sapiens membrane-spanning 4-domains, subfamily A, member 12 (MS4A12), mRNA.; Membrane-spanning 4-domains, subfamily A, member 12
NM_14 5168		NAD(P) dependent steroid dehydrogenase-like; Homo sapiens NAD(P) dependent steroid dehydrogenase-like (HSPC105), mRNA.
NM_00 0371	T TR	transthyretin (prealbumin, amyloidosis type I); synonyms: PALB, TBPA, HsT2651; Homo sapiens transthyretin (prealbumin, amyloidosis type I) (TTR), mRNA.; Transthyretin (prealbumin, amyloidosis type I)
NM_01 4467	S RPX2	sushi-repeat-containing protein, X-linked 2; synonyms: RESDX, SRPUL; sushi-repeat protein; bA524D16A.1 (sushi-repeat-containing protein); Homo sapiens sushi-repeat-containing protein, X-linked 2 (SRPX2), mRNA.; Sushi-repeat-containing protein, X-linked 2
NM_01 9079	L1 TD1	LINE-1 type transposase domain containing 1; synonyms: ECAT11, FLJ10884, MGC133253, RP5-

		1155K23.3; Homo sapiens LINE-1 type transposase domain containing 1 (L1TD1), mRNA.
NM_002763	P ROX1	prospero homeobox 1; prospero-related homeobox 1; Homo sapiens prospero homeobox 1 (PROX1), mRNA.; Prospero-related homeobox 1
NM_018423	S TYK1	serine/threonine/tyrosine kinase 1; synonyms: NOK, SuRTK106, DKFZp761P1010; protein kinase STYK1; Homo sapiens serine/threonine/tyrosine kinase 1 (STYK1), mRNA.; Serine/threonine/tyrosine kinase 1
NM_198148	C PXM2	carboxypeptidase X (M14 family), member 2; synonym: UNQ676; carboxypeptidase-like protein X2; 4632435C11Rik; carboxypeptidase Hlo; Homo sapiens carboxypeptidase X (M14 family), member 2 (CPXM2), mRNA.; Carboxypeptidase X (M14 family), member 2
NM_152321	E RP27	endoplasmic reticulum protein 27 kDa; synonyms: C12orf46, FLJ32115; Homo sapiens endoplasmic reticulum protein 27 kDa (ERP27), mRNA.; Endoplasmic reticulum protein 27 kDa
NM_025176		KIAA0980 protein; synonyms: NLP, FLJ11792, dJ691N24.1; Homo sapiens KIAA0980 protein (KIAA0980), mRNA.
NM_004787	S LIT2	slit homolog 2 (Drosophila); synonyms: SLIL3, Slit-2, FLJ14420; Homo sapiens slit homolog 2 (Drosophila) (SLIT2), mRNA.; Slit homolog 2 (Drosophila)
NM_138809	C MBL	carboxymethylenebutenolidase homolog (Pseudomonas); synonym: FLJ23617; Homo sapiens carboxymethylenebutenolidase homolog (Pseudomonas) (CMBL), mRNA.; Carboxymethylenebutenolidase homolog

		(Pseudomonas)
NM_18 1795	P KIB	protein kinase (cAMP-dependent, catalytic) inhibitor beta; synonyms: PRKACN2, FLJ23817; cAMP-dependent protein kinase inhibitor 2; PKI-beta; Homo sapiens protein kinase (cAMP-dependent, catalytic) inhibitor beta (PKIB), transcript variant 1, mRNA.; Protein kinase (cAMP-dependent, catalytic) inhibitor beta
NM_00 0237	L PL	lipoprotein lipase; synonym: LIPD; Homo sapiens lipoprotein lipase (LPL), mRNA.; Lipoprotein lipase
NM_00 1333	C TSL2	cathepsin L2; synonyms: CTSU, CTSV, CATL2, MGC125957; cathepsin U; cathepsin V; cathepsin L2, preproprotein; Homo sapiens cathepsin L2 (CTSL2), mRNA.; Cathepsin L2
NM_00 2644	PI GR	polymeric immunoglobulin receptor; synonyms: MGC125361, MGC125362; hepatocellular carcinoma associated protein TB6; Homo sapiens polymeric immunoglobulin receptor (PIGR), mRNA.; Polymeric immunoglobulin receptor
NM_00 3063	S LN	sarcolipin; synonyms: MGC12301, MGC125854, MGC125855; Homo sapiens sarcolipin (SLN), mRNA.; Sarcolipin
NM_00 1007139	IG F2	insulin-like growth factor 2 (somatomedin A); synonyms: INSIGF, pp9974, C11orf43, FLJ22066, FLJ44734; insulin-like growth factor type 2; insulin-like growth factor II; putative insulin-like growth factor II associated protein; Homo sapiens insulin-like growth factor 2 (somatomedin A) (IGF2), transcript variant 2, mRNA.; Insulin-like growth factor 2 (somatomedin A)
NM_00	CI	cartilage intermediate layer protein, nucleotide

3613	LP	pyrophosphohydrolase; synonym: HsT18872; Homo sapiens cartilage intermediate layer protein, nucleotide pyrophosphohydrolase (CILP), mRNA.; Cartilage intermediate layer protein, nucleotide pyrophosphohydrolase
ENST0000267857		glucosaminyl (N-acetyl) transferase 3, mucin type [Source:RefSeq_peptide;Acc:NP_004742] [ENST00000267857]
NM_017636	T RPM4	transient receptor potential cation channel, subfamily M, member 4; synonyms: TRPM4B, FLJ20041; Homo sapiens transient receptor potential cation channel, subfamily M, member 4 (TRPM4), mRNA.; Transient receptor potential cation channel, subfamily M, member 4
NM_002193	IN HBB	inhibin, beta B (activin AB beta polypeptide); Inhibin, beta-2; Homo sapiens inhibin, beta B (activin AB beta polypeptide) (INHBB), mRNA.; Inhibin, beta B (activin AB beta polypeptide)
NM_032034	S LC4A11	solute carrier family 4, sodium bicarbonate transporter-like, member 11; synonyms: BTR1, CHED2, NABC1, MGC126418, MGC126419, dJ794I6.2; bicarbonate transporter related protein 1; sodium bicarbonate transporter-like protein 11; corneal endothelial dystrophy 2 (autosomal recessive); Homo sapiens solute carrier family 4, sodium bicarbonate transporter-like, member 11 (SLC4A11), mRNA.; Solute carrier family 4, sodium bicarbonate transporter-like, member 11
NM_032391		small nuclear protein PRAC; synonym: MGC32520; Homo sapiens small nuclear protein PRAC (PRAC), mRNA.; Small nuclear protein PRAC
NM_01	S	sine oculis binding protein homolog



8013	OBP	(Drosophila); synonym: FLJ10159; Homo sapiens sine oculis binding protein homolog (Drosophila) (SOBP), mRNA.; Sine oculis binding protein homolog (Drosophila)
A_23_ P158868		
NM_00 1080494	C 1orf34	chromosome 1 open reading frame 34; synonyms: DEME-6, KIAA0452; Homo sapiens chromosome 1 open reading frame 34 (C1orf34), mRNA.; Chromosome 1 open reading frame 34
NM_02 0997	L EFTY1	left-right determination factor 1; synonyms: LEFTB, LEFTYB; left-right determination, factor B; Homo sapiens left-right determination factor 1 (LEFTY1), mRNA.; Left-right determination factor 1
NM_00 7115	T NFAIP6	tumor necrosis factor, alpha-induced protein 6; synonym: TSG6; tumor necrosis factor-stimulated gene-6 protein; tumor necrosis factor alpha-inducible protein 6; tumor necrosis factor-inducible protein 6; hyaluronate-binding protein; Homo sapiens tumor necrosis factor, alpha-induced protein 6 (TNFAIP6), mRNA.; Tumor necrosis factor, alpha-induced protein 6
NM_14 4646	IG J	immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides; synonyms: JCH, IGCJ; Homo sapiens immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides (IGJ), mRNA.; Immunoglobulin J polypeptide, linker protein for immunoglobulin alpha and mu polypeptides
NM_00 4591	C CL20	chemokine (C-C motif) ligand 20; synonyms: CKb4, LARC, ST38, MIP3A, MIP-3a, SCYA20; macrophage inflammatory protein 3 alpha; small

		inducible cytokine subfamily A (Cys-Cys), member 20; exodus-1; Homo sapiens chemokine (C-C motif) ligand 20 (CCL20), mRNA.; Chemokine (C-C motif) ligand 20
CD709 370		EST25897 human nasopharynx Homo sapiens cDNA, mRNA sequence.; Immunoglobulin V-region of lambda chain
NM_17 0736	K CNJ15	potassium inwardly-rectifying channel, subfamily J, member 15; synonyms: IRKK, KIR1.3, KIR4.2, MGC13584; inward rectifier K <sup>+</sup> channel KIR4.2; ATP-sensitive inward rectifier potassium channel 15; Homo sapiens potassium inwardly-rectifying channel, subfamily J, member 15 (KCNJ15), transcript variant 1, mRNA.; Potassium inwardly-rectifying channel, subfamily J, member 15
NM_00 2416	C XCL9	chemokine (C-X-C motif) ligand 9; synonyms: CMK, MIG, Humig, SCYB9, crg-10; monokine induced by gamma interferon; Homo sapiens chemokine (C-X-C motif) ligand 9 (CXCL9), mRNA.; Chemokine (C-X-C motif) ligand 9
NM_01 6606	R EEP2	receptor accessory protein 2; synonyms: C5orf19, SGC32445; Homo sapiens receptor accessory protein 2 (REEP2), mRNA.; Receptor accessory protein 2
NM_00 5627	S GK	serum/glucocorticoid regulated kinase; synonym: SGK1; serine/threonine protein kinase SGK; serum and glucocorticoid regulated kinase; Homo sapiens serum/glucocorticoid regulated kinase (SGK), mRNA.; Serum/glucocorticoid regulated kinase
NM_01 2101	T RIM29	tripartite motif-containing 29; synonyms: ATDC, FLJ36085; ataxia-telangiectasia group D-

		associated protein; Homo sapiens tripartite motif-containing 29 (TRIM29), mRNA.; Tripartite motif-containing 29
NM_15 2852	M S4A6A	membrane-spanning 4-domains, subfamily A, member 6A; synonyms: CDA01, MS4A6, 4SPAN3, CD20L3, MST090, MSTP090, 4SPAN3.2, MGC22650, MGC131944; isoform 1 is encoded by transcript variant 1; CD20-like precursor; four-span transmembrane protein 3.2; MS4A6A-polymorph; four-span transmembrane protein 3.1; HAIRB-iso; membrane-spanning 4-domains, subfamily A, member 6A, isoform 2; Homo sapiens membrane-spanning 4-domains, subfamily A, member 6A (MS4A6A), transcript variant 1, mRNA.; Membrane-spanning 4-domains, subfamily A, member 6A
NM_00 0518	H BB	hemoglobin, beta; synonyms: HBD, CD113t-C; beta globin chain; Homo sapiens hemoglobin, beta (HBB), mRNA.; Hemoglobin, beta
NM_00 0693	A LDH1A3	aldehyde dehydrogenase 1 family, member A3; synonyms: ALDH6, RALDH3, ALDH1A6; aldehyde dehydrogenase 6; Homo sapiens aldehyde dehydrogenase 1 family, member A3 (ALDH1A3), mRNA.; Aldehyde dehydrogenase 1 family, member A3
NM_01 5957	A PIP	APAF1 interacting protein; synonyms: APIP2, CGI-29, MMRP19, dJ179L10.2; APAF1-interacting protein; Homo sapiens APAF1 interacting protein (APIP), mRNA.
NM_17 4953	A TP2A3	ATPase, Ca++ transporting, ubiquitous; synonym: SERCA3; isoform e is encoded by transcript variant 5; ATPase, Ca(2+)-transporting, ubiquitous; sarcoplasmic/endoplasmic reticulum

		calcium ATPase 3; SR Ca(2+)-ATPase 3; calcium pump 3; adenosine triphosphatase, calcium; sarco/endoplasmic reticulum Ca2+ -ATPase; calcium-translocating P-type ATPase; sarco/endoplasmic reticulum Ca2+ ATPase isoform 3f; Homo sapiens ATPase, Ca++ transporting, ubiquitous (ATP2A3), transcript variant 5, mRNA.
NM_0010104	CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1; synonyms: CP1B, GLC3A; aryl hydrocarbon hydroxylase; microsomal monooxygenase; xenobiotic monooxygenase; flavoprotein-linked monooxygenase; cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1 (glaucoma 3, primary infantile); Homo sapiens cytochrome P450, family 1, subfamily B, polypeptide 1 (CYP1B1), mRNA.; Cytochrome P450, family 1, subfamily B, polypeptide 1
NM_003392	WNT5A	wingless-type MMTV integration site family, member 5A; synonym: hWNT5A; WNT-5A protein; Homo sapiens wingless-type MMTV integration site family, member 5A (WNT5A), mRNA.; Wingless-type MMTV integration site family, member 5A
NM_020208	SLC6A20	solute carrier family 6 (proline IMINO transporter), member 20; synonyms: XT3, SIT1, Xtrp3, MGC161475; isoform 1 is encoded by transcript variant 1; orphan transporter XT3; sodium/imino-acid transporter 1; X transporter protein 3; solute carrier family 6 (neurotransmitter transporter), member 20; neurotransmitter transporter RB21A; Homo sapiens solute carrier family 6 (proline IMINO transporter), member 20 (SLC6A20), transcript variant 1, mRNA.; Solute

		carrier family 6 (proline IMINO transporter), member 20
AB063751		Homo sapiens IGH mRNA for immunoglobulin heavy chain VHDJ region, partial cds, clone:H101.
ENST0000390263		Immunoglobulin Kappa light chain V gene segment [Source:IMG/GENE_DB;Acc:IGKV1-39][ENST00000390263]
NM_001073	UGT2B11	UDP glucuronosyltransferase 2 family, polypeptide B11; synonyms: MGC129611, MGC129612; UDP glycosyltransferase 2 family protein; Homo sapiens UDP glucuronosyltransferase 2 family, polypeptide B11 (UGT2B11), mRNA.
NM_0010253	MTTP	microsomal triglyceride transfer protein; synonyms: ABL, MTP, MGC149819, MGC149820; microsomal triglyceride transfer protein (large polypeptide, 88kD); microsomal triglyceride transfer protein (large polypeptide, 88kDa); Homo sapiens microsomal triglyceride transfer protein (MTTP), mRNA.; Microsomal triglyceride transfer protein
NM_0014887	CXCL14	chemokine (C-X-C motif) ligand 14; synonyms: KS1, Kec, BMAC, BRAK, NJAC, MIP-2g, SCYB14, MGC10687, bolekin; CXC chemokine in breast and kidney; small inducible cytokine subfamily B (Cys-X-Cys), member 14 (BRAK); Homo sapiens chemokine (C-X-C motif) ligand 14 (CXCL14), mRNA.; Chemokine (C-X-C motif) ligand 14
NM_0011964	EGR1	early growth response 1; synonyms: TIS8, AT225, G0S30, NGFI-A, ZNF225, KROX-24, ZIF-268; nerve growth factor-induced protein A; transcription factor ETR103; zinc finger protein 225; early growth response protein 1; Homo sapiens early growth response 1 (EGR1), mRNA.; Early growth

		response 1
NM_15 3840	G PR110	G protein-coupled receptor 110; synonyms: PGR19, KPG_012, hGPCR36, FLJ22684, FLJ30646, MGC125952; isoform 1 is encoded by transcript variant 1; G-protein coupled receptor 110; G protein-coupled receptor PGR19; seven transmembrane helix receptor; Homo sapiens G protein-coupled receptor 110 (GPR110), transcript variant 1, mRNA.
NM_00 3014	S FRP4	secreted frizzled-related protein 4; synonyms: FRP-4, FRPHE, MGC26498; Homo sapiens secreted frizzled-related protein 4 (SFRP4), mRNA.; Secreted frizzled-related protein 4
A_23_ P215382		
NM_21 3674	T PM2	tropomyosin 2 (beta); synonyms: DA1, TMSB, AMCD1; isoform 2 is encoded by transcript variant 2; arthrogryposis multiplex congenital, distal, type 1; Homo sapiens tropomyosin 2 (beta) (TPM2), transcript variant 2, mRNA.; Tropomyosin 2 (beta)
NM_00 4167	C CL15	chemokine (C-C motif) ligand 15; synonyms: LKN1, NCC3, SY15, HCC-2, Lkn-1, MIP-5, NCC-3, SCYL3, MIP-1d, SCYA15, HMRP-2B; small inducible cytokine subfamily A (Cys-Cys), member 15; leukotactin 1; CC chemokine 3; macrophage inflammatory protein 5; chemokine CC-2; MIP-1 delta; Homo sapiens chemokine (C-C motif) ligand 15 (CCL15), transcript variant 2, mRNA.; Chemokine (C-C motif) ligand 15
NM_00 2965	S 100A9	S100 calcium binding protein A9; synonyms: MIF, NIF, P14, CAGB, CFAG, CGLB, L1AG, LIAG, MRP14, 60B8AG, MAC387; S100 calcium-binding protein A9 (calgranulin B); S100 calcium binding

		protein A9 (calgranulin B); Homo sapiens S100 calcium binding protein A9 (S100A9), mRNA.
AK126 354	C 10orf81	Homo sapiens cDNA FLJ44383 fis, clone TRACH3036207.; chromosome 10 open reading frame 81; Chromosome 10 open reading frame 81
NM_03 2251	C CDC88B	coiled-coil domain containing 88B; synonyms: BRLZ, HkRP2, CCDC88, FLJ00354, FLJ37970, DKFZp434G0920; brain leucine zipper protein; Homo sapiens coiled-coil domain containing 88B (CCDC88B), mRNA.; Coiled-coil domain containing 88
NM_15 2315	F AM55A	family with sequence similarity 55, member A; synonym: MGC34290; Homo sapiens family with sequence similarity 55, member A (FAM55A), mRNA.; Family with sequence similarity 55, member D
NM_14 4583	A TP6V1C 2	ATPase, H <sup>+</sup> transporting, lysosomal 42kDa, V1 subunit C2; synonyms: VMA5, ATP6C2; isoform b is encoded by transcript variant 2; V-ATPase C2 subunit; ATPase, H <sup>+</sup> transporting, lysosomal 42kD, V1 subunit C; Homo sapiens ATPase, H <sup>+</sup> transporting, lysosomal 42kDa, V1 subunit C2 (ATP6V1C2), transcript variant 2, mRNA.
NM_00 6998	S CGN	secretagogin, EF-hand calcium binding protein; synonyms: SEGN, CALBL, SECRET, setagin, DJ501N12.8; calbindin-like; Homo sapiens secretagogin, EF-hand calcium binding protein (SCGN), mRNA.; Secretagogin, EF-hand calcium binding protein
NM_00 4796	N RXN3	neurexin 3; synonym: KIAA0743; isoform alpha precursor is encoded by transcript variant alpha; neurexin III; Homo sapiens neurexin 3

		(NRXN3), transcript variant alpha, mRNA.; Neurexin 3
NM_002281	KRT81	keratin 81; synonyms: HB1, Hb-1, KRTHB1, MLN137, ghHkb1, hHAKB2-1; hard keratin, type II, 1; Homo sapiens keratin 81 (KRT81), mRNA.; Keratin 81
NM_0010130	C8orf4	chromosome 8 open reading frame 4; synonyms: TC1, TC-1, hTC-1, MGC22806; human thyroid cancer 1; Homo sapiens chromosome 8 open reading frame 4 (C8orf4), mRNA.; Chromosome 8 open reading frame 4
NM_005123	NR1H4	nuclear receptor subfamily 1, group H, member 4; synonyms: BAR, FXR, HRR1, HRR-1, RIP14, MGC163445; farnesoid X receptor; Homo sapiens nuclear receptor subfamily 1, group H, member 4 (NR1H4), mRNA.; Nuclear receptor subfamily 1, group H, member 4
NM_006149	LGALS4	lectin, galactoside-binding, soluble, 4 (galectin 4); synonym: GAL4; lectin galactoside-binding soluble 4; Homo sapiens lectin, galactoside-binding, soluble, 4 (galectin 4) (LGALS4), mRNA.; Lectin, galactoside-binding, soluble, 4 (galectin 4)
NM_000047	ARSE	arylsulfatase E (chondrodysplasia punctata 1); synonyms: CDPX, CDPX1, CDPXR, MGC163310; Homo sapiens arylsulfatase E (chondrodysplasia punctata 1) (ARSE), mRNA.; Arylsulfatase E (chondrodysplasia punctata 1)
NM_000683	ADRA2C	adrenergic, alpha-2C-, receptor; synonyms: ADRA2L2, ADRARL2, ADRA2RL2, ALPHA2CAR; alpha2-AR-C4; Homo sapiens adrenergic, alpha-2C-, receptor (ADRA2C), mRNA.; Adrenergic, alpha-2C-, receptor



NM_01 4479	A DAMDE C1	ADAM-like, decysin 1; synonym: M12.219; disintegrin protease; Homo sapiens ADAM-like, decysin 1 (ADAMDEC1), mRNA.
NM_00 1873	C PE	carboxypeptidase E; Homo sapiens carboxypeptidase E (CPE), mRNA.; Carboxypeptidase E
AF471 475		Homo sapiens clone 10-45m Ig heavy chain variable region, VH3 family mRNA, partial cds.
NM_00 1169	A QP8	aquaporin 8; Homo sapiens aquaporin 8 (AQP8), mRNA.; Aquaporin 8
NM_01 4859		KIAA0672 gene product; Homo sapiens KIAA0672 gene product (KIAA0672), mRNA.
NM_00 6418	O LFM4	olfactomedin 4; synonyms: GC1, OlfD, GW112, KIAA4294, bA209J19.1; differentially expressed in hematopoietic lineages; bA209J19.1.1 (GW112 protein); Homo sapiens olfactomedin 4 (OLFM4), mRNA.; Olfactomedin 4
NM_08 0672	P HACTR3	phosphatase and actin regulator 3; synonyms: H17739, SCAPIN1, C20orf101, MGC117178; isoform 1 is encoded by transcript variant 1; scapinin; scaffold-associated PP1 inhibiting protein; phosphatase and actin regulator 3, isoform 1; Homo sapiens phosphatase and actin regulator 3 (PHACTR3), transcript variant 1, mRNA.; Phosphatase and actin regulator 3
ENST0 0000251847		Kinesin-like motor protein C20orf23 (Sorting nexin-23). [Source:Uniprot/SWISSPROT;Acc:Q96L93] [ENST00000251847]
NM_02 0685	C 3orf14	chromosome 3 open reading frame 14; synonym: HT021; Homo sapiens chromosome 3 open reading frame 14 (C3orf14), mRNA.;

		Chromosome 3 open reading frame 14
NM_01 3955	N OX1	NADPH oxidase 1; synonyms: MOX1, NOH1, NOH-1, GP91-2; isoform long variant is encoded by transcript variant NOH-1Lv; mitogenic oxidase (pyridine nucleotide-dependent superoxide-generating); NADPH oxidase homolog-1; NADPH oxidase 1 variant NOH-1L; Homo sapiens NADPH oxidase 1 (NOX1), transcript variant NOH-1Lv, mRNA.
NM_13 8340	A BHD3	abhydrolase domain containing 3; synonyms: LABH3, MGC11259; lung alpha/beta hydrolase 3; Homo sapiens abhydrolase domain containing 3 (ABHD3), mRNA.; Abhydrolase domain containing 3
NM_00 6235	P OU2AF1	POU domain, class 2, associating factor 1; synonyms: BOB1, OBF1, OCAB, OBF-1; Homo sapiens POU domain, class 2, associating factor 1 (POU2AF1), mRNA.
NM_01 7781	C YP2W1	cytochrome P450, family 2, subfamily W, polypeptide 1; synonym: MGC34287; Homo sapiens cytochrome P450, family 2, subfamily W, polypeptide 1 (CYP2W1), mRNA.; Cytochrome P450, family 2, subfamily W, polypeptide 1
NM_03 3317	D MKN	dermokine; synonyms: UNQ729, ZD52F10; isoform beta is encoded by transcript variant 2; dermokine beta-1; dermokine beta-2; Homo sapiens dermokine (DMKN), transcript variant 2, mRNA.; Dermokine
NM_00 0300	P LA2G2A	phospholipase A2, group IIA (platelets, synovial fluid); synonyms: MOM1, PLA2, PLA2B, PLA2L, PLA2S, PLAS1, sPLA2; phosphatidylcholine 2-acylhydrolase; synovial phospholipase-A2; Homo sapiens phospholipase A2, group IIA (platelets,

		synovial fluid) (PLA2G2A), mRNA.; Phospholipase A2, group IIA (platelets, synovial fluid)
NM_02 1199	S QRDL	sulfide quinone reductase-like (yeast); synonym: CGI-44; Homo sapiens sulfide quinone reductase-like (yeast) (SQRDL), mRNA.; Sulfide quinone reductase-like (yeast)
NM_00 1039567	R PS4Y2	ribosomal protein S4, Y-linked 2; synonym: RPS4Y2P; ribosomal protein S4, Y-linked 2 pseudogene; Homo sapiens ribosomal protein S4, Y-linked 2 (RPS4Y2), mRNA.; Ribosomal protein S4, Y-linked 2
NM_00 6671	S LC1A7	solute carrier family 1 (glutamate transporter), member 7; synonyms: AAAT, EAAT5, FLJ36602; excitatory amino acid transporter 5 (retinal glutamate transporter); Homo sapiens solute carrier family 1 (glutamate transporter), member 7 (SLC1A7), mRNA.; Solute carrier family 1 (glutamate transporter), member 7
NM_08 0759	D ACH1	dachshund homolog 1 (Drosophila); synonyms: DACH, FLJ10138; isoform a is encoded by transcript variant 1; Homo sapiens dachshund homolog 1 (Drosophila) (DACH1), transcript variant 1, mRNA.; Dachshund homolog 1 (Drosophila)
AK055 387	S PEG	Homo sapiens cDNA FLJ30825 fis, clone FEBRA2001706, highly similar to Human APEG-1 mRNA.; SPEG complex locus
NM_00 1399	E DA	ectodysplasin A; synonyms: ED1, HED, EDA1, EDA2, XHED, XLHED, ED1-A1, ED1-A2; isoform EDA-A1 is encoded by transcript variant 1; ectodysplasin-A; ectodermal dysplasia 1, anhidrotic; ectodermal dysplasia, anhidrotic (hypohydrotic); Homo sapiens ectodysplasin A (EDA), transcript

		variant 1, mRNA.; Ectodysplasin A
NM_17 8537	B 4GALNT 4	beta-1,4-N-acetyl-galactosaminyl transferase 4; synonyms: FLJ25045, NGalNAc-T1, Beta4GalNAc-T4; beta1,4-N-acetylgalactosaminyltransferase III; Homo sapiens beta-1,4-N-acetyl-galactosaminyl transferase 4 (B4GALNT4), mRNA.; Beta-1,4-N-acetyl-galactosaminyl transferase 4
NM_02 2370	R OBO3	roundabout, axon guidance receptor, homolog 3 (Drosophila); synonyms: HGPS, RIG1, HGPPS, RBIG1, FLJ21044; retinoblastoma inhibiting gene 1; horizontal gaze palsy with progressive scoliosis; Homo sapiens roundabout, axon guidance receptor, homolog 3 (Drosophila) (ROBO3), mRNA.; Roundabout, axon guidance receptor, homolog 3 (Drosophila)
NM_02 2073	E GLN3	egl nine homolog 3 (C. elegans); synonyms: PHD3, HIFPH3, FLJ21620, MGC125998, MGC125999; EGL nine (C.elegans) homolog 3; Homo sapiens egl nine homolog 3 (C. elegans) (EGLN3), mRNA.; Egl nine homolog 3 (C. elegans)
BC040 542	KI AA0802	Homo sapiens KIAA0802, mRNA (cDNA clone MGC:39663 IMAGE:5268201), complete cds.
NM_00 5099	A DAMTS4	ADAM metalloproteinase with thrombospondin type 1 motif, 4; synonyms: ADMP-1, ADAMTS-2, ADAMTS-4, KIAA0688; aggrecanase-1; a disintegrin-like and metalloprotease (reprolysin type) with thrombospondin type 1 motif, 4; Homo sapiens ADAM metalloproteinase with thrombospondin type 1 motif, 4 (ADAMTS4), mRNA.
NM_02 2901	L RRC19	leucine rich repeat containing 19; synonym: FLJ21302; Homo sapiens leucine rich repeat

		containing 19 (LRRC19), mRNA.; Leucine rich repeat containing 19
NM_14 4505	K LK8	kallikrein-related peptidase 8; synonyms: NP, HNP, NRPN, PRSS19, TADG14; isoform 2 is encoded by transcript variant 2; protease, serine, 19; neuropsin; ovasin; tumor-associated differentially expressed gene 14; neuropsin type 1; neuropsin type 2; kallikrein 8 (neuropsin/ovasin); Homo sapiens kallikrein-related peptidase 8 (KLK8), transcript variant 2, mRNA.; Kallikrein-related peptidase 8
NM_00 1192	T NFRSF1 7	tumor necrosis factor receptor superfamily, member 17; synonyms: BCM, BCMA, CD269; B-cell maturation factor; Homo sapiens tumor necrosis factor receptor superfamily, member 17 (TNFRSF17), mRNA.; Tumor necrosis factor receptor superfamily, member 17
NM_01 6434	R TEL1	regulator of telomere elongation helicase 1; synonyms: NHL, RTEL, C20orf41, KIAA1088, bK3184A7.3, DKFZP434C013; isoform 1 is encoded by transcript variant 1; regulator of telomere length; Homo sapiens regulator of telomere elongation helicase 1 (RTEL1), transcript variant 1, mRNA.; Tumor necrosis factor receptor superfamily, member 6b, decoy
NM_01 6569	T BX3	T-box 3 (ulnar mammary syndrome); synonyms: UMS, XHL, TBX3-ISO; isoform 2 is encoded by transcript variant 2; bladder cancer related protein XHL; T-box transcription factor TBX3; T-box 3 protein; Homo sapiens T-box 3 (ulnar mammary syndrome) (TBX3), transcript variant 2, mRNA.
NM_15	T	transmembrane protein 92; synonym:

3229	MEM92	FLJ33318; Homo sapiens transmembrane protein 92 (TMEM92), mRNA.; Transmembrane protein 92
NM_05 3277	C LIC6	chloride intracellular channel 6; synonym: CLIC1L; chloride channel form A; Homo sapiens chloride intracellular channel 6 (CLIC6), mRNA.; Chloride intracellular channel 6
NM_01 7763	R NF43	ring finger protein 43; synonyms: URCC, RNF124, FLJ20315, MGC125630, DKFZp781H0392, DKFZp781H02126; Homo sapiens ring finger protein 43 (RNF43), mRNA.; Ring finger protein 43
NM_00 6528	T FPI2	tissue factor pathway inhibitor 2; synonyms: PP5, TFPI-2, FLJ21164; Homo sapiens tissue factor pathway inhibitor 2 (TFPI2), mRNA.; Tissue factor pathway inhibitor 2
NM_00 4335	B ST2	bone marrow stromal cell antigen 2; synonym: CD317; Homo sapiens bone marrow stromal cell antigen 2 (BST2), mRNA.; Bone marrow stromal cell antigen 2
AK027 091		Homo sapiens cDNA: FLJ23438 fis, clone HRC13275.; CDNA: FLJ23438 fis, clone HRC13275
NM_00 1853	C OL9A3	collagen, type IX, alpha 3; synonyms: IDD, MED, EDM3, FLJ90759, DJ885L7.4.1; collagen IX, alpha-3 polypeptide; collagen type IX proteoglycan; Homo sapiens collagen, type IX, alpha 3 (COL9A3), mRNA.; Collagen, type IX, alpha 3
NM_00 5711	E DIL3	EGF-like repeats and discoidin I-like domains 3; synonyms: DEL1, MGC26287; developmental endothelial locus-1; Homo sapiens EGF-like repeats and discoidin I-like domains 3 (EDIL3), mRNA.
NM_01 2261	C 20orf103	chromosome 20 open reading frame 103; Homo sapiens chromosome 20 open reading frame 103 (C20orf103), mRNA.; Chromosome 20 open

		reading frame 103
NM_001236	C BR3	carbonyl reductase 3; synonym: hCBR3; carbonyl reductase (NADPH) 3; NADPH-dependent carbonyl reductase 3; Homo sapiens carbonyl reductase 3 (CBR3), mRNA.; Carbonyl reductase 3
NM_012465	TL L2	tolloid-like 2; synonyms: KIAA0932, MGC133312, MGC133313; Homo sapiens tolloid-like 2 (TLL2), mRNA.; Tolloid-like 2
NM_176782	C 1orf179	chromosome 1 open reading frame 179; synonym: MGC27169; hypothetical protein MGC27169; Homo sapiens chromosome 1 open reading frame 179 (C1orf179), mRNA.; Chromosome 1 open reading frame 179
NM_153026	P RICKLE 1	prickle homolog 1 (Drosophila); synonyms: RILP, FLJ31627, FLJ31937, MGC138902, MGC138903; REST (RE-1 silencing transcription factor)/NRSF (neuron-restrictive silencer factor)-interacting LIM domain protein; REST/NRSF-interacting LIM domain protein; Homo sapiens prickle homolog 1 (Drosophila) (PRICKLE1), mRNA.; Prickle homolog 1 (Drosophila)
NM_005213	C STA	cystatin A (stefin A); synonyms: STF1, STFA; cystatin AS; Homo sapiens cystatin A (stefin A) (CSTA), mRNA.; Cystatin A (stefin A)
NM_000612	IG F2	insulin-like growth factor 2 (somatomedin A); synonyms: INSIGF, pp9974, C11orf43, FLJ22066, FLJ44734; insulin-like growth factor type 2; insulin-like growth factor II; putative insulin-like growth factor II associated protein; Homo sapiens insulin-like growth factor 2 (somatomedin A) (IGF2), transcript variant 1, mRNA.; Insulin-like growth factor 2 (somatomedin A)

NM_003734	AOC3	amine oxidase, copper containing 3 (vascular adhesion protein 1); synonyms: HPAO, SSAO, VAP1, VAP-1; copper amine oxidase; semicarbazide-sensitive amine oxidase; Homo sapiens amine oxidase, copper containing 3 (vascular adhesion protein 1) (AOC3), mRNA.; Amine oxidase, copper containing 3 (vascular adhesion protein 1)
X97261	MT1L	H.sapiens mRNA for metallothionein isoform 1R.; metallothionein 1L (gene/pseudogene); Metallothionein 1L (gene/pseudogene)
NM_153270	KLHL34	kelch-like 34 (Drosophila); synonyms: FLJ34960, MGC125650, RP11-450P7.3; Homo sapiens kelch-like 34 (Drosophila) (KLHL34), mRNA.; Kelch-like 34 (Drosophila)
NM_015170	SULF1	sulfatase 1; synonyms: SULF-1, HSULF-1, FLJ30905, FLJ38022, FLJ41750, KIAA1077; sulfatase FP; Homo sapiens sulfatase 1 (SULF1), mRNA.; Sulfatase 1
NM_178859		organic solute transporter beta; synonyms: FLJ26090, MGC118959, MGC118960, MGC118961; Homo sapiens organic solute transporter beta (OSTbeta), mRNA.; Organic solute transporter beta
NM_001445	FABP6	fatty acid binding protein 6, ileal (gastrotropin); synonyms: ILBP, I-15P, I-BAP, ILBP3, ILLBP, I-BABP, I-BALB; isoform 2 is encoded by transcript variant 2; illeal lipid-binding protein; ileal bile acid binding protein; Homo sapiens fatty acid binding protein 6, ileal (gastrotropin) (FABP6), transcript variant 2, mRNA.; Fatty acid binding protein 6, ileal (gastrotropin)
NM_030949	PP1R14	protein phosphatase 1, regulatory (inhibitor) subunit 14C; synonyms: KEPI, NY-BR-81, CPI17-



	C	like; serologically defined breast cancer antigen NY-BR-81; PKC-potentiated PP1 inhibitory protein; Homo sapiens protein phosphatase 1, regulatory (inhibitor) subunit 14C (PPP1R14C), mRNA.; Protein phosphatase 1, regulatory (inhibitor) subunit 14C
NM_021637	T MEM35	transmembrane protein 35; synonym: FLJ14084; Homo sapiens transmembrane protein 35 (TMEM35), mRNA.; Transmembrane protein 35
NM_000399	E GR2	early growth response 2 (Krox-20 homolog, Drosophila); synonyms: CMT1D, CMT4E, KROX20, FLJ14547, DKFZp686J1957; KROX-20, Drosophila, homolog (early growth response-2); Homo sapiens early growth response 2 (Krox-20 homolog, Drosophila) (EGR2), mRNA.; Early growth response 2 (Krox-20 homolog, Drosophila)
NM_006533	MI A	melanoma inhibitory activity; synonym: CD-RAP; Homo sapiens melanoma inhibitory activity (MIA), mRNA.; Melanoma inhibitory activity
NM_006840	LI LRB5	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 5; synonyms: LIR8, CD85C, LIR-8; isoform 2 is encoded by transcript variant 2; leukocyte immunoglobulin-like receptor 8; Homo sapiens leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 5 (LILRB5), transcript variant 2, mRNA.; Leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 5
NM_032895		hypothetical protein MGC14376; synonyms: MGC149751, DKFZp686O06159; Homo sapiens hypothetical protein MGC14376 (MGC14376), transcript variant 1, mRNA.; Hypothetical protein

		MGC14376
NM_17 2113	E YA2	eyes absent homolog 2 (Drosophila); synonyms: EAB1, MGC10614; isoform b is encoded by transcript variant 2; translation of this uORF probably lowers the translation efficiency of EYA2; Homo sapiens eyes absent homolog 2 (Drosophila) (EYA2), transcript variant 2, mRNA.; Eyes absent homolog 2 (Drosophila)
NM_01 2342	B AMBI	BMP and activin membrane-bound inhibitor homolog (Xenopus laevis); synonym: NMA; non-metastatic gene A protein; putative transmembrane protein; Homo sapiens BMP and activin membrane-bound inhibitor homolog (Xenopus laevis) (BAMBI), mRNA.
NM_00 0494	C OL17A1	collagen, type XVII, alpha 1; synonyms: BP180, BPAG2, LAD-1, KIAA0204, BA16H23.2; collagen XVII, alpha-1 polypeptide; epidermolysis bullosa, junctional, localisata; bullous pemphigoid antigen 2 (180kD); bA16H23.2 (collagen, type XVII, alpha 1 (BP180)); Homo sapiens collagen, type XVII, alpha 1 (COL17A1), mRNA.; Collagen, type XVII, alpha 1
NM_17 0744	U NC5B	unc-5 homolog B (C. elegans); synonyms: UNC5H2, p53RDL1; unc5 (C.elegans homolog) b; transmembrane receptor Unc5H2; p53-regulated receptor for death and life; Homo sapiens unc-5 homolog B (C. elegans) (UNC5B), mRNA.; Unc-5 homolog B (C. elegans)
NM_15 2718	V WCE	von Willebrand factor C and EGF domains; synonyms: VWC1, URG11, FLJ32009; Homo sapiens von Willebrand factor C and EGF domains (VWCE), mRNA.; Von Willebrand factor C and EGF

		domains
NM_01 8690		apolipoprotein B48 receptor; Homo sapiens apolipoprotein B48 receptor (APOB48R), mRNA.; Apolipoprotein B48 receptor
BC107 798	T NNT1	synonym: ANM; Homo sapiens troponin T type 1 (skeletal, slow), mRNA (cDNA clone MGC:104241 IMAGE:4247379), complete cds.; Troponin T type 1 (skeletal, slow)
NM_00 4460	F AP	fibroblast activation protein, alpha; synonyms: FAPA, DPPIV, SEPRASE, DKFZp686G13158; integral membrane serine protease; Homo sapiens fibroblast activation protein, alpha (FAP), mRNA.; Fibroblast activation protein, alpha
NM_00 5940	M MP11	matrix metalloproteinase 11 (stromelysin 3); synonyms: ST3, SL-3, STMY3; matrix metalloproteinase 11 (stromelysin 3); stromelysin III; Homo sapiens matrix metalloproteinase 11 (stromelysin 3) (MMP11), mRNA.; Matrix metalloproteinase 11 (stromelysin 3)
NM_02 1101	C LDN1	claudin 1; synonyms: CLD1, SEMP1, ILVASC; senescence-associated epithelial membrane protein 1; Homo sapiens claudin 1 (CLDN1), mRNA.; Claudin 1
NM_19 9511	C CDC80	coiled-coil domain containing 80; synonyms: URB, DRO1, SSG1, okuribin, MGC131805, MGC134851; steroid sensitive gene 1; nuclear envelope protein okuribin; Homo sapiens coiled-coil domain containing 80 (CCDC80), transcript variant 1, mRNA.; Coiled-coil domain containing 80
NM_00 5980	S 100P	S100 calcium binding protein P; synonym: MIG9; S100 calcium-binding protein P; migration-inducing gene 9; Homo sapiens S100 calcium

		binding protein P (S100P), mRNA.
NM_02 4563	C 5orf23	chromosome 5 open reading frame 23; synonym: FLJ14054; Homo sapiens chromosome 5 open reading frame 23 (C5orf23), mRNA.; Chromosome 5 open reading frame 23
NM_00 1012761	R GMB	RGM domain family, member B; synonyms: DRAGON, FLJ90406, MGC86970; isoform 1 precursor is encoded by transcript variant 1; repulsive guidance molecule B; Homo sapiens RGM domain family, member B (RGMB), transcript variant 1, mRNA.
NM_03 0615	KI F25	kinesin family member 25; synonyms: KNSL3, MGC163361; isoform 1 is encoded by transcript variant 1; kinesin-like 3; Homo sapiens kinesin family member 25 (KIF25), transcript variant 1, mRNA.; Kinesin family member 25
NM_00 5950	M T1G	metallothionein 1G; synonyms: MT1, MT1K, MGC12386; metallothionein 1K; Homo sapiens metallothionein 1G (MT1G), mRNA.; Metallothionein 1G
A_23_ P61068		
NM_01 6848	S HC3	SHC (Src homology 2 domain containing) transforming protein 3; synonyms: NSHC, SHCC, N- Shc; neuronal Shc; src homology 2 domain- containing transforming protein C3; Homo sapiens SHC (Src homology 2 domain containing) transforming protein 3 (SHC3), mRNA.
NM_00 3020	S CG5	secretogranin V (7B2 protein); synonyms: 7B2, SgV, P7B2, SGNE1; prohormone convertase chaperone; Homo sapiens secretogranin V (7B2 protein) (SCG5), mRNA.; Secretogranin V (7B2

		protein)
NM_03 2229	S LITRK6	SLIT and NTRK-like family, member 6; synonyms: MGC119595, MGC119596, MGC119597; 4832410J21Rik; slit and trk like gene 6; Homo sapiens SLIT and NTRK-like family, member 6 (SLITRK6), mRNA.
NM_00 0138	F BN1	fibrillin 1; synonyms: FBN, SGS, WMS, MASS, MFS1, OCTD; fibrillin 1 (Marfan syndrome); fibrillin 15; Homo sapiens fibrillin 1 (FBN1), mRNA.; Fibrillin 1
NM_01 9010	K RT20	keratin 20; synonyms: K20, CD20, CK20, KRT21, MGC35423; cytokeratin 20; keratin, type I cytoskeletal 20; Homo sapiens keratin 20 (KRT20), mRNA.; Keratin 20
NM_03 3050	S UCNR1	succinate receptor 1; synonym: GPR91; G protein-coupled receptor 91; P2Y purinoceptor 1; Homo sapiens succinate receptor 1 (SUCNR1), mRNA.; Succinate receptor 1
NM_00 6681	N MU	neuromedin U; Homo sapiens neuromedin U (NMU), mRNA.; Neuromedin U
NM_02 2138	S MOC2	SPARC related modular calcium binding 2; synonyms: SMAP2, MST117, SMOC-2, MSTP117, MSTP140, bA37D8.1, bA270C4A.1, dJ421D16.1; smooth muscle associated protein 2; dJ421D16.1 (novel thyroglobulin type-1 repeat containing protein); bA37D8.1 (secreted modular calcium-binding protein 2); Homo sapiens SPARC related modular calcium binding 2 (SMOC2), mRNA.
NM_00 5527	H SPA1L	heat shock 70kDa protein 1-like; synonyms: hum70t, HSP70-HOM; heat shock 70kD protein-like 1; heat shock 10kDa protein 1-like; Homo sapiens heat shock 70kDa protein 1-like (HSPA1L), mRNA.;

		Heat shock 70kDa protein 1-like
NM_001511	CXCL1	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha); synonyms: GRO1, GROa, MGSA, NAP-3, SCYB1, MGSA-a, MGSA alpha; GRO1 oncogene (melanoma growth-stimulating activity); GRO1 oncogene (melanoma growth stimulating activity, alpha); melanoma growth stimulatory activity alpha; Homo sapiens chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha) (CXCL1), mRNA.; Chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)
NM_000204	CFI	complement factor I; synonyms: FI, IF, KAF, C3b-INa, factor I; light chain of factor I; I factor (complement); C3b-inactivator; Konglutinogen-activating factor; Homo sapiens complement factor I (CFI), mRNA.; Complement factor I
A_23_P72252		
AY003763		VH3 family, 3-30/cos-3; Homo sapiens isolate sy-3A/17-G9 immunoglobulin alpha heavy chain variable region mRNA, partial cds.; Immunoglobulin heavy constant gamma 1 (G1m marker)
NM_015714	G0S2	G0/G1switch 2; synonym: RP1-28O10.2; putative lymphocyte G0/G1 switch; Homo sapiens G0/G1switch 2 (G0S2), mRNA.
NM_006744	RBP4	retinol binding protein 4, plasma; retinol-binding protein 4, plasma; retinol-binding protein 4, interstitial; Homo sapiens retinol binding protein 4, plasma (RBP4), mRNA.; Retinol binding protein 4, plasma
NM_02	M	membrane-spanning 4-domains, subfamily A,

4021	S4A4A	member 4; synonyms: MS4A4, MS4A7, 4SPAN1, CD20L1, CD20-L1, HDCME31P, MGC22311; isoform a is encoded by transcript variant 1; membrane-spanning 4-domains, subfamily A, member 4A; four-span transmembrane protein; Fc epsilon receptor beta subunit homolog; Homo sapiens membrane-spanning 4-domains, subfamily A, member 4 (MS4A4A), transcript variant 1, mRNA.; Membrane-spanning 4-domains, subfamily A, member 4
NM_01 6584	IL 23A	interleukin 23, alpha subunit p19; synonyms: P19, SGRF, IL-23, IL-23A, IL23P19, MGC79388; interleukin 23 p19 subunit; JKA3 induced upon T-cell activation; Homo sapiens interleukin 23, alpha subunit p19 (IL23A), mRNA.; Interleukin 23, alpha subunit p19
NM_01 4226	R AGE	renal tumor antigen; synonyms: MOK, RAGE1; renal cell carcinoma antigen (MOK protein kinase); Homo sapiens renal tumor antigen (RAGE), mRNA.; Renal tumor antigen
NM_00 0576	IL 1B	interleukin 1, beta; synonyms: IL-1, IL1F2, IL1-BETA; preinterleukin 1 beta; catabolin; pro-interleukin-1-beta; Homo sapiens interleukin 1, beta (IL1B), mRNA.; Interleukin 1, beta
NR_00 1442	F ER1L4	fer-1-like 4 (C. elegans); synonyms: FLJ13459, FLJ22613, C20orf124, dJ309K20.1, bA563A22B.1; Homo sapiens fer-1-like 4 (C. elegans) (FER1L4) on chromosome 20.
NM_00 4164	R BP2	retinol binding protein 2, cellular; synonyms: CRBP2, RBPC2, CRBP2, CRABP-II; retinol-binding protein 2, cellular; Homo sapiens retinol binding protein 2, cellular (RBP2), mRNA.; Retinol binding

		protein 2, cellular
NM_002123	HLA-DQB1	major histocompatibility complex, class II, DQ beta 1; synonyms: IDDM1, CELIAC1, HLA-DQB; MHC class II HLA-DQ beta glycoprotein; MHC class II antigen HLA-DQ-beta-1; MHC DQ beta; MHC class2 antigen; major histocompatibility complex class II beta; lymphocyte antigen; Homo sapiens major histocompatibility complex, class II, DQ beta 1 (HLA-DQB1), mRNA.; Major histocompatibility complex, class II, DQ beta 1
NM_006398	UBD	ubiquitin D; synonyms: FAT10, UBD-3; Homo sapiens ubiquitin D (UBD), mRNA.; Ubiquitin D
NM_002395	ME1	malic enzyme 1, NADP(+)-dependent, cytosolic; synonyms: MES, HUMNDME; pyruvic-malic carboxylase; Malic enzyme, cytoplasmic; malic enzyme 1, soluble; NADP-dependent malic enzyme; malate dehydrogenase; Homo sapiens malic enzyme 1, NADP(+)-dependent, cytosolic (ME1), mRNA.; Malic enzyme 1, NADP(+)-dependent, cytosolic
NM_002523	NPTX2	neuronal pentraxin II; synonyms: NP2, NARP, NP-II; neuronal activity-regulated pentaxin; Homo sapiens neuronal pentraxin II (NPTX2), mRNA.; Neuronal pentraxin II
NM_033014	OGN	osteoglycin; synonyms: OIF, SLRR3A, DKFZP586P2421; preproprotein isoform 2 is encoded by transcript variant 1; osteoinductive factor; osteoglycin (osteoinductive factor); mimecan proteoglycan; Homo sapiens osteoglycin (OGN), transcript variant 1, mRNA.; Osteoglycin
NM_002048	GAS1	growth arrest-specific 1; Growth arrest-specific gene-1; Homo sapiens growth arrest-specific 1 (GAS1), mRNA.; Growth arrest-specific 1



NM_01 6307	P RRX2	paired related homeobox 2; synonyms: PMX2, PRX2, MGC19843; paired-like homeodomain protein PRX2; paired mesoderm homeobox protein 2; predicted NUP98/PRRX2 fusion protein; Homo sapiens paired related homeobox 2 (PRRX2), mRNA.; Paired related homeobox 2
A_23_ P84791		
NM_00 0640	IL 13RA2	interleukin 13 receptor, alpha 2; synonyms: IL-13R, IL13BP, CD213A2; interleukin 13 binding protein; interleukin 13 receptor alpha 2 chain; IL-13 receptor; Homo sapiens interleukin 13 receptor, alpha 2 (IL13RA2), mRNA.; Interleukin 13 receptor, alpha 2
NM_00 3944	S ELENBP 1	selenium binding protein 1; synonyms: LPSB, SP56, hSBP, hSP56, FLJ13813; Homo sapiens selenium binding protein 1 (SELENBP1), mRNA.; Selenium binding protein 1
NM_00 0777	C YP3A5	cytochrome P450, family 3, subfamily A, polypeptide 5; synonyms: CP35, PCN3, P450PCN3; flavoprotein-linked monooxygenase; niphedipine oxidase; cytochrome P450, subfamily IIIA (niphedipine oxidase), polypeptide 5; aryl hydrocarbon hydroxylase; xenobiotic monooxygenase; microsomal monooxygenase; cytochrome P-450; Homo sapiens cytochrome P450, family 3, subfamily A, polypeptide 5 (CYP3A5), mRNA.; Cytochrome P450, family 3, subfamily A, polypeptide 5
NM_00 0067	C A2	carbonic anhydrase II; synonyms: CAII, Car2, CA II, CA-II; carbonate dehydratase II; carbonic dehydratase; carbonic anhydrase B; carbonic

		anhydrase II, CA II; Homo sapiens carbonic anhydrase II (CA2), mRNA.; Carbonic anhydrase II
NM_001927	DES	desmin; synonyms: CSM1, CSM2, CMD1I, FLJ12025, FLJ39719, FLJ41013, FLJ41793; intermediate filament protein; Homo sapiens desmin (DES), mRNA.; Desmin
NM_002345	LUM	lumican; synonyms: LDC, SLRR2D; lumican proteoglycan; Homo sapiens lumican (LUM), mRNA.; Lumican
ENST0000390615		Immunoglobulin heavy chain V gene segment [Source:IMGT/GENE_DB;Acc:IGHV3-33] [ENST00000390615]
A_24_P110487		
NM_003869	CES2	carboxylesterase 2 (intestine, liver); synonyms: iCE, CE-2, PCE-2, CES2A1; isoform 1 is encoded by transcript variant 1; intestinal carboxylesterase; liver carboxylesterase-2; Homo sapiens carboxylesterase 2 (intestine, liver) (CES2), transcript variant 1, mRNA.; Carboxylesterase 2 (intestine, liver)
NM_000517	HBA2	hemoglobin, alpha 2; synonym: HBA1; alpha globin; alpha-2 globin; Homo sapiens hemoglobin, alpha 2 (HBA2), mRNA.; Hemoglobin, alpha 2
Y11328		H.sapiens mRNA for immunoglobulin heavy chain, partial, clone VH3-7.; Immunoglobulin heavy chain, partial, clone VH3-7
AK097639	CBWD5	unnamed protein product; Homo sapiens cDNA FLJ40320 fis, clone TESTI2030770.; COBW domain containing 5
ENST0000372045		Chordin-like protein 1 precursor (Neuralin-1) (Ventroptin) (Neurogenesin-1).

		[Source:Uniprot/SWISSPROT;Acc:Q9BU40] [ENST00000372045]
NM_001040105	MUC17	mucin 17, cell surface associated; synonym: MUC3; small intestinal mucin MUC3; membrane mucin MUC17; secreted mucin MUC17; Homo sapiens mucin 17, cell surface associated (MUC17), mRNA.; Mucin 17, cell surface associated
NM_016619	PLAC8	placenta-specific 8; synonyms: C15, onzin; Homo sapiens placenta-specific 8 (PLAC8), mRNA.; Placenta-specific 8
NM_015393		DKFZP564O0823 protein; Homo sapiens DKFZP564O0823 protein (DKFZP564O0823), mRNA.
NM_002546	TNFRSF11B	tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin); synonyms: OPG, TR1, OCIF, MGC29565; osteoclastogenesis inhibitory factor; Homo sapiens tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin) (TNFRSF11B), mRNA.; Tumor necrosis factor receptor superfamily, member 11b (osteoprotegerin)
ENST0000390258		Immunoglobulin Kappa light chain V gene segment [Source:IMGT/GENE_DB;Acc:IGKV1-27] [ENST00000390258]
NM_0030341	SLC3A1	solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1; synonyms: D2H, ATR1, NBAT, RBAT, CSNU1, FLJ34681; solute carrier family 3 (cystine, dibasic and neutral amino acid transporters), member 1; amino acid transporter 1; SLC3A1 variant F; SLC3A1 variant G; SLC3A1 variant B; SLC3A1 variant C; SLC3A1 variant D; SLC3A1 variant E; Homo sapiens

		solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1 (SLC3A1), mRNA.; Solute carrier family 3 (cystine, dibasic and neutral amino acid transporters, activator of cystine, dibasic and neutral amino acid transport), member 1
NM_006890	C EACAM7	carcinoembryonic antigen-related cell adhesion molecule 7; synonyms: CEA, CGM2; carcinoembryonic antigen gene family member 2; Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 7 (CEACAM7), mRNA.; Carcinoembryonic antigen-related cell adhesion molecule 7
NM_000870	H TR4	5-hydroxytryptamine (serotonin) receptor 4; synonyms: 5-HT4, 5-HT4R; isoform b is encoded by transcript variant b; 5-hydroxytryptamine4 receptor; cardiac 5-HT4 receptor; serotonin 5-HT4 receptor; Homo sapiens 5-hydroxytryptamine (serotonin) receptor 4 (HTR4), transcript variant b, mRNA.
XR_018247		Derived by automated computational analysis using gene prediction method: GNOMON. Supporting evidence includes similarity to: 9 Proteins; PREDICTED: Homo sapiens similar to 40S ribosomal protein S4, Y isoform 1 (LOC441774), mRNA.; Similar to 40S ribosomal protein S4, Y isoform 1
NM_001013618		similar to omega protein; Homo sapiens similar to omega protein (LOC91353), mRNA.; Similar to omega protein
ENST0000390543		Immunoglobulin heavy chain C gene segment [Source:IMG/GENE_DB;Acc:IGHG4] [ENST00000390543]

BC067 092	IG KC	synonyms: HCAK1, Km; Homo sapiens immunoglobulin kappa constant, mRNA (cDNA clone MGC:71265 IMAGE:6302855), complete cds.; Major histocompatibility complex, class I, C
NM_00 2994	C XCL5	chemokine (C-X-C motif) ligand 5; synonyms: SCYB5, ENA-78; small inducible cytokine subfamily B (Cys-X-Cys), member 5 (epithelial-derived neutrophil-activating peptide 78); neutrophil-activating peptide ENA-78; small inducible cytokine B5; epithelial-derived neutrophil activating protein 78; neutrophil-activating protein 78; Homo sapiens chemokine (C-X-C motif) ligand 5 (CXCL5), mRNA.; Chemokine (C-X-C motif) ligand 5
BC103 878	M OGAT2	synonyms: MGAT2, DGAT2L5, FLJ22644; Homo sapiens monoacylglycerol O-acyltransferase 2, mRNA (cDNA clone MGC:119185 IMAGE:40004283), complete cds.; Monoacylglycerol O-acyltransferase 2
NM_03 3342	T TRIM7	tripartite motif-containing 7; synonyms: GNIP, RNF90; isoform 4 is encoded by transcript variant 6; glycogenin-interacting protein; tripartite motif protein TRIM7; Homo sapiens tripartite motif-containing 7 (TRIM7), transcript variant 6, mRNA.; Tripartite motif-containing 7
NM_08 0657	R RSAD2	radical S-adenosyl methionine domain containing 2; synonyms: cig5, vig1, 2510004L01Rik; viperin; Homo sapiens radical S-adenosyl methionine domain containing 2 (RSAD2), mRNA.; Radical S-adenosyl methionine domain containing 2
NM_00 3225	T FF1	trefoil factor 1; synonyms: pS2, BCE1, HPS2, HP1.A, pNR-2, D21S21; breast cancer estrogen-inducible sequence; gastrointestinal trefoil protein

		pS2; Homo sapiens trefoil factor 1 (TFF1), mRNA.; Trefoil factor 1
NM_02 1012	K CNJ12	potassium inwardly-rectifying channel, subfamily J, member 12; synonyms: IRK2, hIRK, hIRK1, KCNJN1, Kir2.2, Kir2.2v, kcnj12x, hkir2.2x; ATP-sensitive inward rectifier potassium channel 12; potassium inwardly-rectifying channel, subfamily J, inhibitor 1; Homo sapiens potassium inwardly-rectifying channel, subfamily J, member 12 (KCNJ12), mRNA.; Potassium inwardly-rectifying channel, subfamily J, member 12
NM_00 6475	P OSTN	periostin, osteoblast specific factor; synonyms: PN, OSF-2, PDLPOSTN, MGC119510, MGC119511, periostin, RP11-412K4.1; osteoblast specific factor 2 (fasciclin I-like); periodontal ligament-specific periostin; Homo sapiens periostin, osteoblast specific factor (POSTN), mRNA.; Periostin, osteoblast specific factor
NM_00 4496	F OXA1	forkhead box A1; synonyms: HNF3A, TCF3A, MGC33105; hepatocyte nuclear factor 3, alpha; Homo sapiens forkhead box A1 (FOXA1), mRNA.; Forkhead box A1
NM_00 1926	D EFA6	defensin, alpha 6, Paneth cell-specific; synonyms: DEF6, HD-6; defensin 6; Homo sapiens defensin, alpha 6, Paneth cell-specific (DEFA6), mRNA.; Defensin, alpha 6, Paneth cell-specific
NM_00 6361	H OXB13	homeobox B13; synonym: PSGD; homeo box B13; Homo sapiens homeobox B13 (HOXB13), mRNA.; Homeobox B13
NM_00 7231	S LC6A14	solute carrier family 6 (amino acid transporter), member 14; synonym: ATB(0+); amino acid transporter B0+; solute carrier family 6

		(neurotransmitter transporter), member 14; Homo sapiens solute carrier family 6 (amino acid transporter), member 14 (SLC6A14), mRNA.; Solute carrier family 6 (amino acid transporter), member 14
NM_00565	IL6R	interleukin 6 receptor; synonyms: CD126, IL6RA, IL-6R-1, MGC104991, IL-6R-alpha; isoform 1 precursor is encoded by transcript variant 1; CD126 antigen; interleukin 6 receptor alpha subunit; Homo sapiens interleukin 6 receptor (IL6R), transcript variant 1, mRNA.; Interleukin 6 receptor
NM_001712	CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein); synonyms: BGP, BGP1, BGPI; isoform 1 precursor is encoded by transcript variant 1; antigen CD66; CD66a antigen; biliary glycoprotein adhesion molecule; Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1), transcript variant 1, mRNA.; Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)
NM_0012672		organic solute transporter alpha; synonym: MGC39807; Homo sapiens organic solute transporter alpha (OSTalpha), mRNA.; Organic solute transporter alpha
NM_0010818	OXGR1	oxoglutarate (alpha-ketoglutarate) receptor 1; synonyms: GPR80, GPR99, P2Y15, P2RY15, MGC119206, MGC119207, MGC119208; G protein-coupled receptor 99; seven transmembrane helix receptor; P2Y-like nucleotide receptor; G protein-coupled receptor 80; Homo sapiens oxoglutarate (alpha-ketoglutarate) receptor 1 (OXGR1), mRNA.; Oxoglutarate (alpha-ketoglutarate) receptor 1

NM_002776	KLK10	kallikrein-related peptidase 10; synonyms: NES1, PRSSL1; protease, serine-like, 1; normal epithelial cell-specific 1; breast normal epithelial cell associated serine protease; kallikrein 10; Homo sapiens kallikrein-related peptidase 10 (KLK10), transcript variant 1, mRNA.; Kallikrein-related peptidase 10
NM_001001396	ATP2B4	ATPase, Ca++ transporting, plasma membrane 4; synonyms: MXRA1, PMCA4, ATP2B2, DKFZp686M088, DKFZp686G08106; isoform 4a is encoded by transcript variant 1; matrix-remodelling associated 1; sarcolemmal calcium pump; plasma membrane calcium pump; Homo sapiens ATPase, Ca++ transporting, plasma membrane 4 (ATP2B4), transcript variant 1, mRNA.
NM_0018569	GPR126	G protein-coupled receptor 126; synonyms: DREG, VIGR, PS1TP2; beta 1 is encoded by transcript variant b1; developmentally regulated G-protein-coupled receptor; HBV PreS1-transactivated protein 2; Homo sapiens G protein-coupled receptor 126 (GPR126), transcript variant b1, mRNA.
A_024_P413286		
ENST00000390559		Immunoglobulin heavy chain C gene segment [Source:IMGT/GENE_DB;Acc:IGHM] [ENST00000390559]
NM_0012941	GBP4	guanylate binding protein 4; synonym: Mpa2; Homo sapiens guanylate binding protein 4 (GBP4), mRNA.; Guanylate binding protein 4
NM_0013043	SLC6A6	solute carrier family 6 (neurotransmitter transporter, taurine), member 6; synonyms: TAUT, MGC10619, MGC131729; Homo sapiens solute



		carrier family 6 (neurotransmitter transporter, taurine), member 6 (SLC6A6), mRNA.; Solute carrier family 6 (neurotransmitter transporter, taurine), member 6
NM_003004	S ECTM1	secreted and transmembrane 1; synonym: K12; type 1a transmembrane protein; K12 protein; Homo sapiens secreted and transmembrane 1 (SECTM1), mRNA.; Secreted and transmembrane 1
NM_181789	G LDN	gliomedin; synonyms: CLOM, COLM, CRGL2, CRG-L2, UNC-112, FLJ23917; collomin; colmedin; Homo sapiens gliomedin (GLDN), mRNA.; Gliomedin
ENST0000390275		Immunoglobulin Kappa light chain V gene segment [Source:IMGT/GENE_DB;Acc:IGKV1D-13] [ENST00000390275]
NR_002196	H 19	H19, imprinted maternally expressed untranslated mRNA; synonyms: ASM, BWS, ASM1, MGC4485, PRO2605, D11S813E; Homo sapiens H19, imprinted maternally expressed untranslated mRNA (H19) on chromosome 11.
NM_182495	F AM55B	family with sequence similarity 55, member B; synonym: FLJ25224; Homo sapiens family with sequence similarity 55, member B (FAM55B), mRNA.; Family with sequence similarity 55, member B
NM_001013642		hypothetical LOC388610; Homo sapiens hypothetical LOC388610 (LOC388610), mRNA.
BX640887		Homo sapiens mRNA; cDNA DKFZp686B0328 (from clone DKFZp686B0328).; hypothetical LOC388796; Hypothetical LOC388796
AF471454		Homo sapiens clone 68-46a Ig heavy chain variable region, VH3 family mRNA, partial cds.
NM_00	H	hydroxyprostaglandin dehydrogenase 15-

0860	PGD	(NAD); synonyms: PGDH, PGDH1, 15-PGDH; 15-hydroxyprostaglandin dehydrogenase; NAD+-dependent 15-hydroxyprostaglandin dehydrogenase; 15-hydroxy prostaglandin dehydrogenase; Homo sapiens hydroxyprostaglandin dehydrogenase 15-(NAD) (HPGD), mRNA.; Hydroxyprostaglandin dehydrogenase 15-(NAD)
AF348 994	M T1JP	Homo sapiens MTB (MTB) mRNA, complete cds.; metallothionein 1J (pseudogene); Metallothionein 1F
X0114 7		unnamed protein product; variable region (put. NEW protein) (aa 1-110); Human mRNA for immunoglobulin lambda variable region corresponding to NEW protein of V lambda subgroup I.; Anti-thyroglobulin light chain variable region
THC25 82296		1EA8_A Chain A, Apolipoprotein E3 22kd Fragment Lys146glu Mutant. {Homo sapiens} (exp=-1; wgp=-1; cg=-1), partial (64%) [THC2582296]
A_24_ P798431		
NM_02 0070	IG LL1	immunoglobulin lambda-like polypeptide 1; synonyms: IGO, 14.1, IGL1, IGL5, IGLL, IGVPB, CD179b, VPREB2, IGLJ14.1; isoform a precursor is encoded by transcript variant 1; lambda5; CD179b antigen; immunoglobulin omega polypeptide chain; immunoglobulin-related 14.1 protein; Pre-B lymphocyte-specific protein-2; Homo sapiens immunoglobulin lambda-like polypeptide 1 (IGLL1), transcript variant 1, mRNA.; Immunoglobulin lambda-like polypeptide 1
AK024 680		Homo sapiens cDNA: FLJ21027 fis, clone CAE07110.; CDNA: FLJ21027 fis, clone CAE07110

NM_01 5687	FI LIP1	filamin A interacting protein 1; synonyms: FILIP, KIAA1275; Homo sapiens filamin A interacting protein 1 (FILIP1), mRNA.; Filamin A interacting protein 1
NM_01 8043	T MEM16A	transmembrane protein 16A; synonyms: TAOS2, ORAOV2, FLJ10261; oral cancer overexpressed 2; transmembrane protein 16A (eight membrane-spanning domains); tumor amplified and overexpressed sequence 2; Homo sapiens transmembrane protein 16A (TMEM16A), mRNA.; Transmembrane protein 16A
AL833 309		Homo sapiens mRNA; cDNA DKFZp313A137 (from clone DKFZp313A137).; MRNA
NM_00 0849	G STM3	glutathione S-transferase M3 (brain); synonyms: GST5, GSTB, GTM3, GSTM3-3, MGC3310, MGC3704; glutathione S-transferase, Mu-3; brain GST; glutathione S-alkyltransferase M3; glutathione S-aryltransferase M3; S-(hydroxyalkyl)glutathione lyase M3; glutathione S-aralkyltransferase M3; GST class-mu 3; Homo sapiens glutathione S-transferase M3 (brain) (GSTM3), mRNA.; Glutathione S-transferase M3 (brain)
A_24_ P930327		
NM_00 1874	C PM	carboxypeptidase M; Homo sapiens carboxypeptidase M (CPM), transcript variant 1, mRNA.; Carboxypeptidase M
NM_02 1935	P ROK2	prokineticin 2; synonyms: BV8, PK2, KAL4, MIT1; Homo sapiens prokineticin 2 (PROK2), mRNA.; Prokineticin 2
A_32_		

P112263		
ENST0000370548		Heparan sulfate 2-O-sulfotransferase 1 (EC 2.8.2.-) (2-O-sulfotransferase) (2OST). [Source:Uniprot/SWISSPROT;Acc:Q7LGA3] [ENST00000370548]
NM_198151		hypothetical protein LOC253012; isoform 2 is encoded by transcript variant 2; Homo sapiens hypothetical protein LOC253012 (LOC253012), transcript variant 2, mRNA.; Hypothetical protein LOC253012
THC2664215		
NM_015668	RGS22	regulator of G-protein signalling 22; synonyms: FLJ40080, PRTD-NY2, MGC102908, DKFZP434I092; Homo sapiens regulator of G-protein signalling 22 (RGS22), mRNA.; Regulator of G-protein signalling 22
AA417913		zv94a06.s1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:767410 3', mRNA sequence.; Dicer1, Dcr-1 homolog (Drosophila)
NM_000129	F13A1	coagulation factor XIII, A1 polypeptide; synonym: F13A; TGase; Coagulation factor XIII, A polypeptide; factor XIIIa; bA525O21.1 (coagulation factor XIII, A1 polypeptide); Coagulation factor XIII A chain; Homo sapiens coagulation factor XIII, A1 polypeptide (F13A1), mRNA.; Coagulation factor XIII, A1 polypeptide
ENST0000261275		Uncharacterized protein KIAA0574 (Fragment). [Source:Uniprot/SWISSPROT;Acc:O60320] [ENST00000261275]
ENST0		HUMIGKPB Ig kappa chain {Homo sapiens}

0000331696		(exp=-1; wgp=0; cg=0), complete [THC2557512]
BC095 489	IG KC	synonyms: HCAK1, Km; Homo sapiens immunoglobulin kappa constant, mRNA (cDNA clone MGC:111575 IMAGE:30328747), complete cds.; Major histocompatibility complex, class I, C
THC26 75062		
A_32_ P166653		
THC27 06212		
XM_00 1133802		Derived by automated computational analysis using gene prediction method: GNOMON. Supporting evidence includes similarity to: 16 ESTs; PREDICTED: Homo sapiens hypothetical protein LOC729464 (LOC729464), mRNA.; Hypothetical protein LOC729464
THC26 78045		
ENST0 0000390622		Immunoglobulin heavy chain V gene segment [Source:IMG/GENE_DB;Acc:IGHV1-46] [ENST00000390622]
BM129 308		if20d02.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens cDNA clone IMAGE:5677082 3', mRNA sequence.; Transcribed locus
DQ249 310		cancer-specific; Homo sapiens clone DU2 UCA1 protein (UCA1) mRNA, complete cds.; urothelial cancer associated 1; Urothelial cancer associated 1
THC25 32455		

ENST0000367013		Homo sapiens cDNA clone IMAGE:30520069. [BC094794]
AJ276555		ORF1; Homo sapiens mRNA for hypothetical protein (ORF1), clone 00275.; Hypothetical protein (ORF1), clone 00275
AK093351		unnamed protein product; Homo sapiens cDNA FLJ36032 fis, clone TESTI2017069.; hypothetical protein FLJ36032; Hypothetical protein FLJ36032
AK129844		Homo sapiens cDNA FLJ26334 fis, clone HRT02648.; Phospholamban
ENST0000360524		CDNA FLJ43983 fis, clone TESTI4018881. [Source:Uniprot/SPTREMBL;Acc:Q6ZU57] [ENST00000360524]
NM_005559	L AMA1	laminin, alpha 1; synonym: LAMA; Homo sapiens laminin, alpha 1 (LAMA1), mRNA.; Laminin, alpha 1
NM_003759	S LC4A4	solute carrier family 4, sodium bicarbonate cotransporter, member 4; synonyms: KNBC, NBC1, NBC2, pNBC, HNBC1, hhNMC, SLC4A5, DKFZp781H1314; isoform 2 is encoded by transcript variant 2; sodium bicarbonate cotransporter 1 (sodium bicarbonate cotransporter, kidney; sodium bicarbonate cotransporter, pancreas); solute carrier family 4, sodium bicarbonate cotransporter, member 5; Homo sapiens solute carrier family 4, sodium bicarbonate cotransporter, member 4 (SLC4A4), transcript variant 2, mRNA.; Solute carrier family 4, sodium bicarbonate cotransporter, member 4
THC2655811		
NM_14	C	chromosome 1 open reading frame 125;

4696	1orf125	synonyms: FLJ25438, FLJ32940, DKFZp686H1423, RP11-215I23.2; isoform 1 is encoded by transcript variant 1; Homo sapiens chromosome 1 open reading frame 125 (C1orf125), transcript variant 1, mRNA.; Chromosome 1 open reading frame 125
THC26 57737		NKG2D_HUMAN (P26718) NKG2-D type II integral membrane protein (NKG2-D-activating NK receptor) (NK cell receptor D) (Killer cell lectin-like receptor subfamily K member 1) (CD314 antigen), partial (14%) [THC2657737]
BC064 586		Homo sapiens hypothetical protein LOC145837, mRNA (cDNA clone IMAGE:5740791), partial cds.; Hypothetical protein LOC145837
THC26 47658		
AK074 614	IG F2	unnamed protein product; Homo sapiens cDNA FLJ90133 fis, clone HEMBB1000567.; insulin-like growth factor 2 (somatomedin A)
ENST0 0000327926		Homo sapiens isolate donor Z clone Z1K immunoglobulin kappa light chain variable region mRNA, partial cds. [AF103555]
NM_00 1008387	R EG3G	regenerating islet-derived 3 gamma; synonyms: PAP1B, PAPIB, UNQ429, REG-III, MGC118998, MGC118999, MGC119001; regenerating gene III; LPPM429; PAP IB protein; Homo sapiens regenerating islet-derived 3 gamma (REG3G), transcript variant 1, mRNA.; Regenerating islet-derived 3 gamma
THC26 52817		

### 3.2. GENES SELECCIONADOS EN ESTE TRABAJO NO COMUNES AL ESTUDIO DEL GRUPO DE LA DRA. PÉREZ-VILLAMIL

La siguiente tabla muestra en la primera columna el identificador del gen empleado en el análisis. En la segunda columna el símbolo del gen que corresponde con el identificador y, por último, en la tercera columna una breve descripción del gen.

De los 215 genes que no comparten ambos estudios, propios, de esta tesis, 21 no están aún clasificados por las bases de datos de genes. Estos genes no tienen información en la columna *Descripción*.

Identificador	Símbolo	Descripción
NM_005248	FGFR	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog; synonyms: SRC2, c-fgr, c-src2, FLJ43153, MGC75096, p55c-fgr, p58c-fgr; c-fgr protooncogene; p55-c-fgr protein; c-src-2 protooncogene; Homo sapiens Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog (FGR), transcript variant 1, mRNA.
NM_018171	APPL2	adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 2; synonyms: DIP13B, FLJ10659; adaptor protein containing PH domain, PTB domain and leucine zipper motif 2; Homo sapiens adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 2 (APPL2), mRNA.; Adaptor protein, phosphotyrosine interaction, PH domain and leucine zipper containing 2
NM_20	FH	four and a half LIM domains 2; synonyms:



1555	L2	DRAL, AAG11, SLIM3; down-regulated in rhabdomyosarcoma LIM protein; aging-associated gene 11; four and a half LIM-domain protein 2; Homo sapiens four and a half LIM domains 2 (FHL2), transcript variant 2, mRNA.; Four and a half LIM domains 2
NM_01 4008	CC DC22	coiled-coil domain containing 22; synonyms: JM1, CXorf37; Homo sapiens coiled-coil domain containing 22 (CCDC22), mRNA.; Coiled-coil domain containing 22
NM_00 6406	PR DX4	peroxiredoxin 4; synonym: AOE37-2; thioredoxin peroxidase (antioxidant enzyme); Homo sapiens peroxiredoxin 4 (PRDX4), mRNA.; Peroxiredoxin 4
NM_14 6421	GS TM1	glutathione S-transferase M1; synonyms: MU, H-B, GST1, GTH4, GTM1, MU-1, GSTM1-1, MGC26563, GSTM1a-1a, GSTM1b-1b; isoform 2 is encoded by transcript variant 2; HB subunit 4; glutathione S-transferase, Mu-1; S-(hydroxyalkyl)glutathione lyase; GST class-mu 1; glutathione S-alkyltransferase; glutathione S-aryltransferase; glutathione S-aralkyltransferase; Homo sapiens glutathione S-transferase M1 (GSTM1), transcript variant 2, mRNA.; Glutathione S-transferase M1
NM_01 5149	RG L1	ral guanine nucleotide dissociation stimulator-like 1; synonyms: RGL, KIAA0959; RalGDS-like; Homo sapiens ral guanine nucleotide dissociation stimulator-like 1 (RGL1), mRNA.; Ral guanine nucleotide dissociation stimulator-like 1
NM_01 7550	MIE R2	mesoderm induction early response 1, family member 2; synonym: KIAA1193; Homo sapiens mesoderm induction early response 1, family member 2 (MIER2), mRNA.; Mesoderm induction early response 1, family member 2
NM_00 4173	SL C7A4	solute carrier family 7 (cationic amino acid transporter, y+ system), member 4; synonyms: VH, CAT4, CAT-4, HCAT3, MGC129976, MGC129977; Ig heavy chain variable region; VH 3 family; Homo sapiens solute carrier family 7 (cationic amino acid transporter, y+ system), member 4 (SLC7A4), mRNA.; Solute carrier family 7 (cationic amino acid

		transporter, y+ system), member 4
NM_007054	KIF3A	kinesin family member 3A; kinesin family protein 3A; Homo sapiens kinesin family member 3A (KIF3A), mRNA.; Kinesin family member 3A
NM_002069	GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1; synonym: Gi; Gi1 protein alpha subunit; Homo sapiens guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1 (GNAI1), mRNA.; Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1
U88048		ORF; highly expressed in metastasis-suppressed chromosome 6 melanoma hybrids; Human clone KiSS-16 unknown product mRNA, complete cds.
NM_178834	LAYN	layilin; synonyms: FLJ30977, FLJ31092; Homo sapiens layilin (LAYN), mRNA.; Layilin
NM_032829	C12orf34	chromosome 12 open reading frame 34; synonym: FLJ14721; Homo sapiens chromosome 12 open reading frame 34 (C12orf34), mRNA.; Chromosome 12 open reading frame 34
NM_152270	SLFN11	schlafen family member 11; synonyms: SLFN8/9, FLJ34922; Homo sapiens schlafen family member 11 (SLFN11), mRNA.; Schlafen family member 11
NM_018645	HES6	hairy and enhancer of split 6 (Drosophila); Homo sapiens hairy and enhancer of split 6 (Drosophila) (HES6), mRNA.; Hairy and enhancer of split 6 (Drosophila)
NM_018934	PCDHB14	protocadherin beta 14; synonyms: MGC120065, PCDH-BETA14; Homo sapiens protocadherin beta 14 (PCDHB14), mRNA.; Protocadherin beta 14
NM_032350		hypothetical protein MGC11257; Homo sapiens hypothetical protein MGC11257 (MGC11257), mRNA.; Hypothetical protein MGC11257
NM_017443	POLE3	polymerase (DNA directed), epsilon 3 (p17 subunit); synonyms: p17, YBL1, CHRAC17, CHARAC17; histone fold protein CHRAC17; DNA polymerase epsilon p17 subunit; chromatin

		accessibility complex 17; arsenic transactivated protein; Homo sapiens polymerase (DNA directed), epsilon 3 (p17 subunit) (POLE3), mRNA.; Polymerase (DNA directed), epsilon 3 (p17 subunit)
NM_01 8146	RN MTL1	RNA methyltransferase like 1; synonyms: HC90, FLJ10581; putative RNA methyltransferase; Homo sapiens RNA methyltransferase like 1 (RNMTL1), mRNA.
NM_00 6261	PR OP1	prophet of Pit1, paired-like homeodomain transcription factor; Homo sapiens prophet of Pit1, paired-like homeodomain transcription factor (PROP1), mRNA.; Prophet of Pit1, paired-like homeodomain transcription factor
NM_02 2567	NY X	nyctalopin; synonyms: CLRP, CSNB1, CSNB4, MGC138447; Congenital stationary night blindness-1 (CSNB, complete); congenital stationary night blindness 4; Homo sapiens nyctalopin (NYX), mRNA.; Nyctalopin
NM_17 7967	UB AC2	UBA domain containing 2; synonyms: PHGDHL1, FLJ26351, FLJ30001, FLJ30548, FLJ42413, MGC90487; phosphoglycerate dehydrogenase like 1; RP11-178C10.1; Homo sapiens UBA domain containing 2 (UBAC2), mRNA.
NM_00 2915	RF C3	replication factor C (activator 1) 3, 38kDa; synonyms: RFC38, MGC5276; isoform 1 is encoded by transcript variant 1; replication factor C (activator 1) 3 (38kD); RFC, 38 kD subunit; A1 38 kDa subunit; Homo sapiens replication factor C (activator 1) 3, 38kDa (RFC3), transcript variant 1, mRNA.; Replication factor C (activator 1) 3, 38kDa
BC047 030	UP P1	synonyms: UP, UPASE, UPP, UDRPASE; Homo sapiens uridine phosphorylase 1, mRNA (cDNA clone MGC:54255 IMAGE:5549432), complete cds.; Uridine phosphorylase 1
NR_00 2605	DL EU1	deleted in lymphocytic leukemia, 1; synonyms: BCMS, DLB1, LEU1, XTP6, MGC22430; Homo sapiens deleted in lymphocytic leukemia, 1 (DLEU1) on chromosome 13.
NM_13 0760	MA DCAM1	mucosal vascular addressin cell adhesion molecule 1; synonym: MACAM1; isoform a

		precursor is encoded by transcript variant 1; mucosal addressin cell adhesion molecule-1; Homo sapiens mucosal vascular addressin cell adhesion molecule 1 (MADCAM1), transcript variant 1, mRNA.; Mucosal vascular addressin cell adhesion molecule 1
NM_000292	PHKA2	phosphorylase kinase, alpha 2 (liver); synonyms: PHK, PYK, XLG, PYKL, XLG2; Phosphorylase kinase deficiency, liver (glycogen storage disease; phosphorylase kinase, alpha 2 (liver), glycogen storage disease IX; phosphorylase kinase alpha-subunit; dJ499B10.2 (phosphorylase kinase, alpha 2 (liver) (PYK)); Homo sapiens phosphorylase kinase, alpha 2 (liver) (PHKA2), mRNA.; Phosphorylase kinase, alpha 2 (liver)
BC007360		Homo sapiens hypothetical protein MGC16121, mRNA (cDNA clone IMAGE:3627113), complete cds.; Hypothetical protein MGC16121
NM_004425	ECM1	extracellular matrix protein 1; isoform 1 precursor is encoded by transcript variant 1; secretory component p85; Homo sapiens extracellular matrix protein 1 (ECM1), transcript variant 1, mRNA.; Extracellular matrix protein 1
NM_021969	NR0B2	nuclear receptor subfamily 0, group B, member 2; synonyms: SHP, SHP1; orphan nuclear receptor SHP; small heterodimer partner; Homo sapiens nuclear receptor subfamily 0, group B, member 2 (NR0B2), mRNA.; Nuclear receptor subfamily 0, group B, member 2
NM_032040	CCDC8	coiled-coil domain containing 8; synonym: DKFZP564K0322; Homo sapiens coiled-coil domain containing 8 (CCDC8), mRNA.; Coiled-coil domain containing 8
NM_006286	TFDP2	transcription factor Dp-2 (E2F dimerization partner 2); synonyms: DP2, Dp-2; Homo sapiens transcription factor Dp-2 (E2F dimerization partner 2) (TFDP2), mRNA.; Transcription factor Dp-2 (E2F dimerization partner 2)
NM_003359	UGDH	UDP-glucose dehydrogenase; synonyms: GDH, UGD, UDPGDH, UDP-GlcDH; UDP-glucose 6-dehydrogenase; uridine diphospho-glucose

		dehydrogenase; Homo sapiens UDP-glucose dehydrogenase (UGDH), mRNA.
NM_000497	CYP11B1	cytochrome P450, family 11, subfamily B, polypeptide 1; synonyms: FHI, CPN1, CYP11B, P450C11, FLJ36771, DKFZp686B05283; isoform 1 precursor is encoded by transcript variant 1; steroid 11-beta-monooxygenase; cytochrome p450 XIB1; cytochrome P450, subfamily XIB (steroid 11-beta-hydroxylase), polypeptide 1; Homo sapiens cytochrome P450, family 11, subfamily B, polypeptide 1 (CYP11B1), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.; Cytochrome P450, family 11, subfamily B, polypeptide 1
NR_002767		hypothetical protein PRO2266; Homo sapiens hypothetical protein PRO2266 (PRO2266) on chromosome 6.
NM_020125	SLAMF8	SLAM family member 8; synonyms: BLAME, SBBI42, FLJ20442, MGC129578; BCM-like membrane protein; Homo sapiens SLAM family member 8 (SLAMF8), mRNA.
NM_015917	GSTK1	glutathione S-transferase kappa 1; synonym: GST13; similar to rGSTK1-1; glutathione S-transferase subunit 13 homolog; Homo sapiens glutathione S-transferase kappa 1 (GSTK1), mRNA.; Glutathione S-transferase kappa 1
NM_006265	RAD21	RAD21 homolog (S. pombe); synonyms: HR21, MCD1, NXP1, SCC1, hHR21, HRAD21, FLJ25655, FLJ40596, KIAA0078; protein involved in DNA double-strand break repair; nuclear matrix protein 1; Homo sapiens RAD21 homolog (S. pombe) (RAD21), mRNA.
NM_014732	KIAA0513	KIAA0513; Homo sapiens KIAA0513 (KIAA0513), mRNA.
NM_030579	CYB5B	cytochrome b5 type B (outer mitochondrial membrane); synonyms: OMB5, CYB5-M, DKFZp686M0619; type 2 cyt-b5; outer mitochondrial membrane cytochrome b5; Homo sapiens cytochrome b5 type B (outer mitochondrial membrane) (CYB5B), mRNA.; Cytochrome b5 type B (outer mitochondrial membrane)
NM_00	HP	haptoglobin; synonyms: MGC111141, hp2-

5143		alpha; Homo sapiens haptoglobin (HP), mRNA.; Haptoglobin
NM_02 4881	SL C35E1	solute carrier family 35, member E1; synonyms: FLJ14251, MGC44954; Homo sapiens solute carrier family 35, member E1 (SLC35E1), mRNA.; Solute carrier family 35, member E1
NM_01 3445	GA D1	glutamate decarboxylase 1 (brain, 67kDa); synonyms: GAD, SCP, FLJ45882; isoform GAD25 is encoded by transcript variant GAD25; glutamate decarboxylase 1 (brain, 67kD); Homo sapiens glutamate decarboxylase 1 (brain, 67kDa) (GAD1), transcript variant GAD25, mRNA.; Glutamate decarboxylase 1 (brain, 67kDa)
NM_00 1457	FL NB	filamin B, beta (actin binding protein 278); synonyms: AOI, FH1, SCT, TAP, LRS1, TABP, FLN1L, ABP-278, filamin B, DKFZp686O033, DKFZp686A1668; beta filamin; filamin 1 (actin-binding protein-280)-like; filamin B, beta (actin-binding protein-278); filamin B variant 2; Homo sapiens filamin B, beta (actin binding protein 278) (FLNB), mRNA.; Filamin B, beta (actin binding protein 278)
NM_00 2670	PL S1	plastin 1 (I isoform); synonym: I-PLASTIN; Plastin-1; Homo sapiens plastin 1 (I isoform) (PLS1), mRNA.; Plastin 1 (I isoform)
NM_18 2925	FLT 4	fms-related tyrosine kinase 4; synonyms: PCL, FLT41, VEGFR3; isoform 1 is encoded by transcript variant 1; vascular endothelial growth factor receptor 3; Homo sapiens fms-related tyrosine kinase 4 (FLT4), transcript variant 1, mRNA.; Fms-related tyrosine kinase 4
NM_00 1080849		similar to CG12379-PA; synonym: MGC87491; Homo sapiens similar to CG12379-PA (LOC728489), mRNA.; Similar to CG12379-PA
NM_03 2409	PIN K1	PTEN induced putative kinase 1; synonyms: BRPK, PARK6, FLJ27236; protein kinase BRPK; serine/threonine-protein kinase PINK1; Homo sapiens PTEN induced putative kinase 1 (PINK1), nuclear gene encoding mitochondrial protein, mRNA.
NM_03 2129	PL EKHN1	pleckstrin homology domain containing, family N member 1; synonyms: MGC120613,



		MGC120616, RP11-54O7.7, DKFZP434H2010; Homo sapiens pleckstrin homology domain containing, family N member 1 (PLEKHN1), mRNA.; Pleckstrin homology domain containing, family N member 1
NM_001748	CA PN2	calpain 2, (m/II) large subunit; synonyms: mCANP, CANPL2, CANPml, FLJ39928; calpain, large polypeptide L2; calcium-activated neutral proteinase; Calpain 2, large [catalytic] subunit; Homo sapiens calpain 2, (m/II) large subunit (CAPN2), mRNA.; Calpain 2, (m/II) large subunit
NM_007075	WD R45	WD repeat domain 45; synonyms: JM5, WDRX1, WIPI4, WIPI-4; isoform 1 is encoded by transcript variant 1; WD repeat domain, X-linked 1; WD45 repeat protein interacting with phosphoinositides 4; Homo sapiens WD repeat domain 45 (WDR45), transcript variant 1, mRNA.
BC019234	ZN F273	synonyms: HZF9, MGC12518; Homo sapiens zinc finger protein 273, mRNA (cDNA clone IMAGE:4045234), complete cds.; Zinc finger protein 273
NM_017918	CC DC109B	coiled-coil domain containing 109B; synonym: FLJ20647; Homo sapiens coiled-coil domain containing 109B (CCDC109B), mRNA.; Coiled-coil domain containing 109B
NM_003841	TN FRSF10C	tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain; synonyms: LIT, DCR1, TRID, CD263, TRAILR3, MGC149501, MGC149502; decoy receptor 1; TRAIL receptor 3; TNF related TRAIL receptor; lymphocyte inhibitor of TRAIL; antagonist decoy receptor for TRAIL/Apo-2L; TNF related apoptosis-inducing ligand receptor 3; Homo sapiens tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain (TNFRSF10C), mRNA.; Tumor necrosis factor receptor superfamily, member 10c, decoy without an intracellular domain
NM_020750	XP O5	exportin 5; synonyms: FLJ14239, FLJ32057, FLJ45606, KIAA1291; Homo sapiens exportin 5 (XPO5), mRNA.; Exportin 5
NM_01	RB	RNA binding motif protein 35A; synonym:

7697	M35A	FLJ20171; isoform 1 is encoded by transcript variant 1; Homo sapiens RNA binding motif protein 35A (RBM35A), transcript variant 1, mRNA.
NM_003761	VA MP8	vesicle-associated membrane protein 8 (endobrevin); synonyms: EDB, VAMP5; Homo sapiens vesicle-associated membrane protein 8 (endobrevin) (VAMP8), mRNA.; Vesicle-associated membrane protein 8 (endobrevin)
NM_000859	HM GCR	3-hydroxy-3-methylglutaryl-Coenzyme A reductase; Homo sapiens 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (HMGCR), mRNA.
NM_144994	AN KRD23	ankyrin repeat domain 23; synonyms: DARP, MARP3, FLJ32449, MGC129593; muscle ankyrin repeat protein 3; Homo sapiens ankyrin repeat domain 23 (ANKRD23), mRNA.; Ankyrin repeat domain 23
NM_152326	AN KRD9	ankyrin repeat domain 9; synonym: MGC21990; Homo sapiens ankyrin repeat domain 9 (ANKRD9), mRNA.; Ankyrin repeat domain 9
NM_152424	FA M123B	family with sequence similarity 123B; synonyms: WTX, FLJ39827, RP11-403E24.2; Homo sapiens family with sequence similarity 123B (FAM123B), mRNA.; Family with sequence similarity 123B
NM_019601	SU SD2	sushi domain containing 2; synonyms: BK65A6.2, FLJ22778; Sushi domain (SCR repeat) containing; Homo sapiens sushi domain containing 2 (SUSD2), mRNA.; Sushi domain containing 2
NM_021174	KIA A1967	KIAA1967; synonyms: DBC1, DBC-1; deleted in breast cancer 1; Homo sapiens KIAA1967 (KIAA1967), transcript variant 1, mRNA.
NM_000837	GRI NA	glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding); synonyms: HNRGW, TMBIM3, NMDARA1, MGC99687; transmembrane BAX inhibitor motif containing 3; glutamate receptor, NMDA subtype, glutamate-binding subunit; N-methyl-D-aspartate receptor-associated protein 1; Homo sapiens glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding) (GRINA), transcript variant 1, mRNA.; Glutamate



		receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (glutamate binding)
AF0908 84		Homo sapiens clone HQ0038 PRO0038 mRNA, complete cds.; Clone HQ0038 PRO0038
AF1474 46	LPI N1	Homo sapiens full length insert cDNA clone ZM50A09.; lipin 1; Lipin 1
AK0549 35	TT C9B	unnamed protein product; Homo sapiens cDNA FLJ30373 fis, clone BRACE2007882, weakly similar to Actin-depolymerizing protein N-WASP.; tetratricopeptide repeat domain 9B
NM_00 4135	IDH 3G	isocitrate dehydrogenase 3 (NAD+) gamma; synonym: H-IDHG; isoform a precursor is encoded by transcript variant 1; isocitric dehydrogenase; IDH-gamma; isocitrate dehydrogenase, NAD(+)-specific, mitochondrial, gamma subunit; NAD+-specific ICDH; NAD (H)-specific isocitrate dehydrogenase gamma subunit; Homo sapiens isocitrate dehydrogenase 3 (NAD+) gamma (IDH3G), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.; Isocitrate dehydrogenase 3 (NAD+) gamma
NM_19 8969	AE S	amino-terminal enhancer of split; synonyms: GRG, ESP1, GRG5, TLE5, AES-1, AES-2; isoform a is encoded by transcript variant 1; Homo sapiens amino-terminal enhancer of split (AES), transcript variant 1, mRNA.; Amino-terminal enhancer of split
NM_00 1002901	FC RLB	Fc receptor-like B; synonyms: FcRY, FCRL2, FCRLY, FCRLM2, FREB-2, RP11-474I16.6; Fc receptor like 2; Homo sapiens Fc receptor-like B (FCRLB), mRNA.
NM_00 3882	WI SP1	WNT1 inducible signaling pathway protein 1; synonyms: CCN4, WISP1c, WISP1i, WISP1tc; isoform 1 precursor is encoded by transcript variant 1; wnt-1 signaling pathway protein 1; Wnt-1 inducible signaling pathway protein 1; WNT1 induced secreted protein 1; Homo sapiens WNT1 inducible signaling pathway protein 1 (WISP1), transcript variant 1, mRNA.
NM_00 3562	SL C25A11	solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11; synonyms: OGC, SLC20A4; solute carrier family 20 (oxoglutarate carrier), member 4; Homo sapiens

		solute carrier family 25 (mitochondrial carrier; oxoglutarate carrier), member 11 (SLC25A11), mRNA.; Solute carrier family 25 (mitochondrial carrier
NM_20 7121	FA M110A	family with sequence similarity 110, member A; synonyms: F10, MGC2450, MGC4675, C20orf55, bA371L19.3; Homo sapiens family with sequence similarity 110, member A (FAM110A), transcript variant 2, mRNA.; Family with sequence similarity 110, member A
NM_01 8420	SL C22A15	solute carrier family 22 (organic cation transporter), member 15; synonyms: FLIPT1, PRO34686, DKFZp761G0313; fly-like putative organic ion transporter 1; trans-like protein; Homo sapiens solute carrier family 22 (organic cation transporter), member 15 (SLC22A15), mRNA.; Solute carrier family 22 (organic cation transporter), member 15
NM_00 0312	PR OC	protein C (inactivator of coagulation factors Va and VIIIa); synonym: PROC1; Homo sapiens protein C (inactivator of coagulation factors Va and VIIIa) (PROC), mRNA.; Protein C (inactivator of coagulation factors Va and VIIIa)
NM_00 1080480	MB OAT1	membrane bound O-acyltransferase domain containing 1; synonyms: OACT1, MGC44669, dJ434O11.1; Homo sapiens membrane bound O-acyltransferase domain containing 1 (MBOAT1), mRNA.; Membrane bound O-acyltransferase domain containing 1
NM_15 2705	PO LR1D	polymerase (RNA) I polypeptide D, 16kDa; synonyms: RPA9, RPA16, RPAC2, POLR1C, RPO1-3, MGC9850, FLJ20616; isoform 2 is encoded by transcript variant 2; RNA polymerase I 16 kDa subunit; Homo sapiens polymerase (RNA) I polypeptide D, 16kDa (POLR1D), transcript variant 2, mRNA.; Polymerase (RNA) I polypeptide D, 16kDa
NM_14 4963	FA M91A1	family with sequence similarity 91, member A1; synonyms: FLJ23790, DKFZp666B104; skeletal muscle cells re-entry induced; Homo sapiens family with sequence similarity 91, member A1 (FAM91A1), mRNA.; Family with

		sequence similarity 91, member A1
NM_000579	CCR5	chemokine (C-C motif) receptor 5; synonyms: CKR5, CD195, CKR-5, CCCR5, CMKBR5, CC-CKR-5; chemr13; chemokine (C-C) receptor 5; C-C chemokine receptor 5; chemokine receptor CCR5; CCR5 chemokine receptor; C-C chemokine receptor type 5; Homo sapiens chemokine (C-C motif) receptor 5 (CCR5), mRNA.; Chemokine (C-C motif) receptor 5
NM_017785	CCDC99	coiled-coil domain containing 99; synonym: FLJ20364; Homo sapiens coiled-coil domain containing 99 (CCDC99), mRNA.; Coiled-coil domain containing 99
NM_001024912	CEACAM1	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein); synonyms: BGP, BGP1, BGPI; isoform 2 precursor is encoded by transcript variant 2; antigen CD66; CD66a antigen; biliary glycoprotein adhesion molecule; Homo sapiens carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein) (CEACAM1), transcript variant 2, mRNA.; Carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)
NM_177529	SULT1A1	sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1; synonyms: PST, STP, STP1, P-PST, ST1A3, TSPST1, MGC5163, MGC131921, HAST1/HAST2; isoform a is encoded by transcript variant 2; phenol-preferring phenol sulfotransferase1; phenol-sulfating phenol sulfotransferase; aryl sulfotransferase; thermostable phenol sulfotransferase1; Homo sapiens sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1 (SULT1A1), transcript variant 2, mRNA.; Sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1
ENST0000372821		Polyadenylate-binding protein 1-like. [Source:Uniprot/SWISSPROT;Acc:Q4VXU2] [ENST00000372821]
NM_201542	MEF8	mediator of RNA polymerase II transcription, subunit 8 homolog (S. cerevisiae); synonyms: ARC32, MGC17544, MGC19641; isoform 1 is encoded by transcript variant 1; mediator of RNA

		polymerase II transcription subunit MED8; activator-recruited cofactor 32 kDa component; Homo sapiens mediator of RNA polymerase II transcription, subunit 8 homolog (S. cerevisiae) (MED8), transcript variant 1, mRNA.; Mediator of RNA polymerase II transcription, subunit 8 homolog (S. cerevisiae)
BG108 194		602280063F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4367799 5', mRNA sequence.; Hypoxia-inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)
NM_01 4214	IMP A2	inositol(myo)-1(or 4)-monophosphatase 2; inosine monophosphatase 2; Homo sapiens inositol(myo)-1(or 4)-monophosphatase 2 (IMPA2), mRNA.; Inositol(myo)-1(or 4)-monophosphatase 2
NM_13 9162	SM CR7	Smith-Magenis syndrome chromosome region, candidate 7; synonym: MGC23130; Homo sapiens Smith-Magenis syndrome chromosome region, candidate 7 (SMCR7), transcript variant 1, mRNA.
NM_02 4833	ZN F671	zinc finger protein 671; synonym: FLJ23506; Homo sapiens zinc finger protein 671 (ZNF671), mRNA.; Zinc finger protein 671
NM_01 6368		myo-inositol 1-phosphate synthase A1; Homo sapiens myo-inositol 1-phosphate synthase A1 (ISYNA1), mRNA.; Myo-inositol 1-phosphate synthase A1
NM_02 0435	GJ A12	gap junction protein, alpha 12, 47kDa; synonyms: Cx47, CX46.6, PMLDAR, MGC105119; connexin 47; Homo sapiens gap junction protein, alpha 12, 47kDa (GJA12), mRNA.; Gap junction protein, alpha 12, 47kDa
NM_14 5906		Homo sapiens RIO kinase 3 (yeast) (RIOK3), transcript variant 2, mRNA [NM_145906]
NM_00 1462	FP RL1	formyl peptide receptor-like 1; synonyms: ALXR, HM63, FMLPX, FPR2A, FPRH1, FPRH2, LXA4R, FMLP-R-II; lipoxin A4 receptor (formyl peptide receptor related); Homo sapiens formyl peptide receptor-like 1 (FPRL1), transcript variant 1, mRNA.; Formyl peptide receptor-like 1
NM_00 1250	CD 40	CD40 molecule, TNF receptor superfamily member 5; synonyms: p50, Bp50, CDW40,

		MGC9013, TNFRSF5; isoform 1 precursor is encoded by transcript variant 1; tumor necrosis factor receptor superfamily, member 5; CD40L receptor; CD40 type II isoform; nerve growth factor receptor-related B-lymphocyte activation molecule; B cell-associated molecule; B cell surface antigen CD40; CD40 antigen (TNF receptor superfamily member 5); Homo sapiens CD40 molecule, TNF receptor superfamily member 5 (CD40), transcript variant 1, mRNA.
NM_006670	TP BG	trophoblast glycoprotein; synonyms: 5T4, M6P1, 5T4-AG; 5T4-antigen; Homo sapiens trophoblast glycoprotein (TPBG), mRNA.; Trophoblast glycoprotein
NM_030895	ZN F696	zinc finger protein 696; synonym: FLJ14129; Homo sapiens zinc finger protein 696 (ZNF696), mRNA.; Zinc finger protein 696
NM_024816	RA BEP2	rabaptin, RAB GTPase binding effector protein 2; synonyms: FRA, FLJ23282; Fos-related antigen; Homo sapiens rabaptin, RAB GTPase binding effector protein 2 (RABEP2), mRNA.; Rabaptin, RAB GTPase binding effector protein 2
NM_006760	UP K2	uroplakin 2; synonym: MGC138598; uroplakin II; Homo sapiens uroplakin 2 (UPK2), mRNA.; Uroplakin 2
NM_005694	CO X17	COX17 cytochrome c oxidase assembly homolog (S. cerevisiae); synonyms: MGC104397, MGC117386; human homolog of yeast mitochondrial copper recruitment; Homo sapiens COX17 cytochrome c oxidase assembly homolog (S. cerevisiae) (COX17), nuclear gene encoding mitochondrial protein, mRNA.
NM_002818	PS ME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta); synonyms: PA28B, REGbeta, PA28beta; proteasome activator hPA28 subunit beta; proteasome activator 28-beta; 11S regulator complex beta subunit; MCP activator, 31-kD subunit; activator of multicatalytic protease subunit 2; cell migration-inducing protein 22; Homo sapiens proteasome (prosome, macropain) activator subunit 2 (PA28 beta) (PSME2), mRNA.; Proteasome (prosome, macropain) activator

		subunit 2 (PA28 beta)
NM_14 4683	DH RS13	dehydrogenase/reductase (SDR family) member 13; synonym: MGC23280; Homo sapiens dehydrogenase/reductase (SDR family) member 13 (DHRS13), mRNA.
NM_01 7947	MO COS	molybdenum cofactor sulfurase; synonyms: MOS, HMCS, FLJ20733; Homo sapiens molybdenum cofactor sulfurase (MOCOS), mRNA.; Molybdenum cofactor sulfurase
NM_00 3958	RN F8	ring finger protein 8; synonyms: FLJ12013, KIAA0646; isoform 1 is encoded by transcript variant 1; C3HC4-type zinc finger protein; ring finger protein (C3HC4 type) 8; UBC13/UEV-interacting ring finger protein; Homo sapiens ring finger protein 8 (RNF8), transcript variant 1, mRNA.; Ring finger protein 8
NM_01 7528	WB SCR22	Williams Beuren syndrome chromosome region 22; synonyms: WBMT, PP3381, HUSSY-3, MGC2022, MGC5140, HASJ4442, MGC19709; Williams-Beuren candidate region putative methyltransferase; Homo sapiens Williams Beuren syndrome chromosome region 22 (WBSCR22), mRNA.
NM_00 3828	MT MR1	myotubularin related protein 1; Homo sapiens myotubularin related protein 1 (MTMR1), mRNA.; Myotubularin related protein 1
NM_00 1001965	OR 4D5	olfactory receptor, family 4, subfamily D, member 5; synonym: OR11-276; Homo sapiens olfactory receptor, family 4, subfamily D, member 5 (OR4D5), mRNA.; Olfactory receptor, family 4, subfamily D, member 5
NM_00 1003927	EVI 2A	ecotropic viral integration site 2A; synonyms: EVDA, EVI2; isoform 1 is encoded by transcript variant 1; Homo sapiens ecotropic viral integration site 2A (EVI2A), transcript variant 1, mRNA.; Ecotropic viral integration site 2A
NM_01 4521	SH 3BP4	SH3-domain binding protein 4; synonyms: TTP, BOG25; transferrin receptor trafficking protein; EH-binding protein; Homo sapiens SH3-domain binding protein 4 (SH3BP4), mRNA.
NM_18 1342	FK BP7	FK506 binding protein 7; synonyms: FKBP23, PPIase, MGC9420; peptidyl-prolyl cis-



		trans isomerase; rotamase; FK506-binding protein 7; Homo sapiens FK506 binding protein 7 (FKBP7), mRNA.
NM_15 3638	PA NK2	pantothenate kinase 2 (Hallervorden-Spatz syndrome); synonyms: HSS, HARP, PKAN, NBIA1, C20orf48, MGC15053; isoform 1 preproprotein is encoded by transcript variant 1; neurodegeneration with brain iron accumulation 1 (Hallervorden-Spatz syndrome); pantothenic acid kinase; Homo sapiens pantothenate kinase 2 (Hallervorden-Spatz syndrome) (PANK2), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.; Pantothenate kinase 2 (Hallervorden-Spatz syndrome)
BC007 354		Homo sapiens hypothetical protein MGC16075, mRNA (cDNA clone MGC:16075 IMAGE:3616854), complete cds.; Hypothetical protein MGC16075
NM_00 4762	PS CD1	pleckstrin homology, Sec7 and coiled-coil domains 1(cytohesin 1); synonyms: B2-1, SEC7, D17S811E, FLJ34050, FLJ41900, CYTOHESIN-1; isoform 1 is encoded by transcript variant 1; homolog of secretory protein SEC7; cytoadhesin 1; pleckstrin homology, Sec7 and coiled/coil domains 1(cytohesin 1); Homo sapiens pleckstrin homology, Sec7 and coiled-coil domains 1(cytohesin 1) (PSCD1), transcript variant 1, mRNA.; Pleckstrin homology, Sec7 and coiled-coil domains 1(cytohesin 1)
NM_00 6310	NP EPPS	aminopeptidase puromycin sensitive; synonyms: PSA, MP100; puromycin-sensitive aminopeptidase; metalloproteinase MP100; Homo sapiens aminopeptidase puromycin sensitive (NPEPPS), mRNA.; Aminopeptidase puromycin sensitive
NM_00 4521	KIF 5B	kinesin family member 5B; synonyms: KNS, KINH, KNS1, UKHC; kinesin 1 (110-120kD); kinesin heavy chain; Homo sapiens kinesin family member 5B (KIF5B), mRNA.; Kinesin family member 5B
NM_00 6145	DN AJB1	DnaJ (Hsp40) homolog, subfamily B, member 1; synonyms: Hdj1, HSPF1, Hsp40; heat

		shock 40kD protein 1; DnaJ (Hsp40) homolog, subfamily B, member 1; Homo sapiens DnaJ (Hsp40) homolog, subfamily B, member 1 (DNAJB1), mRNA.
NM_006062	SMYD5	SMYD family member 5; synonyms: RRG1, RAI15, NN8-4AG, ZMYND23; retinoic acid responsive; retinoic acid induced 15; Homo sapiens SMYD family member 5 (SMYD5), mRNA.
ENST0000379019		Putative glycerophosphodiester phosphodiesterase 5 (EC 3.1.-.-). [Source:Uniprot/SWISSPROT;Acc:Q9NPB8] [ENST00000379019]
NM_021971	GMPPB	GDP-mannose pyrophosphorylase B; synonym: KIAA1851; isoform 2 is encoded by transcript variant 2; mannose-1-phosphate guanylyltransferase; Homo sapiens GDP-mannose pyrophosphorylase B (GMPPB), transcript variant 2, mRNA.
NM_016032	ZDHHC9	zinc finger, DHHC-type containing 9; synonyms: DHHC9, CGI-89, ZNF379, ZNF380, CXorf11, ZDHHC10; Asp-His-His-Cys domain containing protein 9; zinc finger, DHHC-type containing 10; zinc finger, DHHC domain containing 10; Homo sapiens zinc finger, DHHC-type containing 9 (ZDHHC9), transcript variant 1, mRNA.; Zinc finger, DHHC-type containing 9
NM_004265	FADS2	fatty acid desaturase 2; synonyms: D6D, DES6, TU13, FADSD6, LLCDL2, SLL0262; delta-6 fatty acid desaturase; linoleoyl-CoA desaturase (delta-6-desaturase)-like 2; Homo sapiens fatty acid desaturase 2 (FADS2), mRNA.; Fatty acid desaturase 2
NM_003291	TPP2	tripeptidyl peptidase II; synonym: FLJ40359; Homo sapiens tripeptidyl peptidase II (TPP2), mRNA.; Tripeptidyl peptidase II
NM_001558	IL10RA	interleukin 10 receptor, alpha; synonyms: IL10R, CDW210A, HIL-10R, IL-10R1; Homo sapiens interleukin 10 receptor, alpha (IL10RA), mRNA.; Interleukin 10 receptor, alpha
NM_002309	LIF	leukemia inhibitory factor (cholinergic differentiation factor); synonyms: CDF, HILDA, D-FACTOR; Homo sapiens leukemia inhibitory factor



		(cholinergic differentiation factor) (LIF), mRNA.; Leukemia inhibitory factor (cholinergic differentiation factor)
NM_17 2171	CA MK2G	calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma; synonyms: CAMK, CAMKG, CAMK-II, FLJ16043, MGC26678; isoform 1 is encoded by transcript variant 1; CaM kinase II; Homo sapiens calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma (CAMK2G), transcript variant 1, mRNA.; Calcium/calmodulin-dependent protein kinase (CaM kinase) II gamma
ENST0 0000329627		Peroxisome assembly protein 26 (Peroxin-26). [Source:Uniprot/SWISSPROT;Acc:Q7Z412] [ENST00000329627]
AK1275 69		unnamed protein product; Homo sapiens cDNA FLJ45662 fis, clone CTONG2027150.; Similar to Reticulocalbin-1 precursor
NM_00 1013723		Homo sapiens hypothetical gene supported by AK094370 (LOC441208), mRNA [NM_001013723]
NM_00 3028	SH B	Src homology 2 domain containing adaptor protein B; synonyms: bA3J10.2, RP11-3J10.8; SHB (Src homology 2 domain containing) adaptor protein B; SHB adaptor protein (a Src homology 2 protein); Homo sapiens Src homology 2 domain containing adaptor protein B (SHB), mRNA.
A_24_P 169593		
XR_01 9103		Derived by automated computational analysis using gene prediction method: GNOMON. Supporting evidence includes similarity to: 1 EST, 5 Proteins; PREDICTED: Homo sapiens similar to Voltage-dependent anion-selective channel protein 3 (VDAC-3) (hVDAC3) (Outer mitochondrial membrane protein porin 3) (LOC341965), mRNA.; Similar to Voltage-dependent anion-selective channel protein 3 (VDAC-3) (hVDAC3) (Outer mitochondrial membrane protein porin 3)
NM_00 1957	ED NRA	endothelin receptor type A; synonyms: ETA, ETRA; G protein-coupled receptor; Homo sapiens endothelin receptor type A (EDNRA), mRNA.; Endothelin receptor type A

NM_01 4953	DIS 3	DIS3 mitotic control homolog (S. cerevisiae); synonyms: RRP44, dis3p, FLJ10484, KIAA1008, MGC33035, bA555G22.1, RP11-342J4.3, DKFZp667L1817; mitotic control protein dis3 homolog; 2810028N01Rik; Homo sapiens DIS3 mitotic control homolog (S. cerevisiae) (DIS3), mRNA.
NM_00 1001437	CC L3L3	chemokine (C-C motif) ligand 3-like 3; synonyms: LD78, 464.2, SCYA3L, SCYA3L1, D17S1718, LD78BETA, MGC12815; chemokine (C-C motif) ligand 3-like, centromeric; small inducible cytokine A3-like 1; G0S19-2; Homo sapiens chemokine (C-C motif) ligand 3-like 3 (CCL3L3), mRNA.; Chemokine (C-C motif) ligand 3-like 3
NM_00 1001937	AT P5A1	ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle; synonyms: OMR, ORM, ATPM, ATP5A, hATP1, ATP5AL2; mitochondrial ATP synthetase, oligomycin-resistant; ATP sythase (F1-ATPase) alpha subunit; ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, alpha subunit, isoform 2, non-cardiac muscle-like 2; ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle; ATP synthase alpha chain, mitochondrial; Homo sapiens ATP synthase, H <sup>+</sup> transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle (ATP5A1), nuclear gene encoding mitochondrial protein, transcript variant 1, mRNA.
XR_01 9115		Derived by automated computational analysis using gene prediction method: GNOMON. Supporting evidence includes similarity to: 3 ESTs, 14 Proteins; PREDICTED: Homo sapiens similar to Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single-strand binding protein) (hnRNP core protein A1) (HDP-1) (Topoisomerase-inhibitor suppressed) (LOC344741), mRNA.; Similar to Heterogeneous nuclear ribonucleoprotein A1 (Helix-destabilizing protein) (Single-strand binding protein) (hnRNP core protein A1) (HDP-1) (Topoisomerase-inhibitor

		suppressed)
NM_006482	DYRK2	dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2; isoform 2 isoform 2 is encoded by transcript variant 2; Homo sapiens dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2 (DYRK2), transcript variant 2, mRNA.; Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 2
NM_020394	ZN F695	zinc finger protein 695; synonyms: SBZF3, MGC61907, RP11-551G24.1; Homo sapiens zinc finger protein 695 (ZNF695), mRNA.; Zinc finger protein 670
NM_199334	THRA	thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian); synonyms: AR7, EAR7, ERBA, ERBA1, NR1A1, THRA1, THRA2, ERB-T-1, MGC43240, c-ERBA-1, MGC000261; isoform 1 is encoded by transcript variant 1; avian erythroblastic leukemia viral (v-erb-a) oncogene homolog; ERBA-related 7; triiodothyronine receptor; Homo sapiens thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian) (THRA), transcript variant 1, mRNA.; Thyroid hormone receptor, alpha (erythroblastic leukemia viral (v-erb-a) oncogene homolog, avian)
A_24_P272910		
NM_139266	STAT1	signal transducer and activator of transcription 1, 91kDa; synonyms: ISGF-3, STAT91, DKFZp686B04100; isoform beta is encoded by transcript variant beta; signal transducer and activator of transcription-1; transcription factor ISGF-3; transcription factor ISGF-3 components p91/p84; Homo sapiens signal transducer and activator of transcription 1, 91kDa (STAT1), transcript variant beta, mRNA.; Signal transducer and activator of transcription 1, 91kDa
NM_017520		M-phase phosphoprotein, mpp8; synonym: FLJ35237; Homo sapiens M-phase phosphoprotein, mpp8 (HSMPP8), mRNA.
NM_03	RA	Ras association (RalGDS/AF-6) domain

2023	SSF4	family 4; synonyms: AD037, MGC44914; tumor suppressor RASSF4; Homo sapiens Ras association (RalGDS/AF-6) domain family 4 (RASSF4), mRNA.
A_24_P 306788		
NM_01 6347	NA T8B	N-acetyltransferase 8B (gene/pseudogene); synonyms: CML2, Hcml2, NAT8BP, MGC97061; N-acetyltransferase Camello 2; Homo sapiens N-acetyltransferase 8B (gene/pseudogene) (NAT8B), mRNA.
A_24_P 315644		
NM_01 4884	SF RS14	splicing factor, arginine/serine-rich 14; synonym: DKFZp779L2418; arginine/serine-rich 14 splicing factor; Homo sapiens splicing factor, arginine/serine-rich 14 (SFRS14), transcript variant 2, mRNA.; Splicing factor, arginine/serine-rich 14
NM_00 4760	ST K17A	serine/threonine kinase 17a; synonym: DRAK1; death-associated protein kinase-related 1; DAP kinase-related apoptosis-inducing protein kinase 1; serine/threonine kinase 17a (apoptosis-inducing); Homo sapiens serine/threonine kinase 17a (STK17A), mRNA.; Serine/threonine kinase 17a
NM_01 4965	TR AK1	trafficking protein, kinesin binding 1; synonym: OIP106; isoform 2 is encoded by transcript variant 2; OGT(O-Glc-NAc transferase)-interacting protein 106 KDa; OGT(O-Glc-NAc transferase) interacting protein 106 KDa; Homo sapiens trafficking protein, kinesin binding 1 (TRAK1), transcript variant 2, mRNA.; Trafficking protein, kinesin binding 1
NM_03 2888	CO L27A1	collagen, type XXVII, alpha 1; synonyms: FLJ11895, KIAA1870, MGC11337, RP11-8211.1; Homo sapiens collagen, type XXVII, alpha 1 (COL27A1), mRNA.; Collagen, type XXVII, alpha 1
NM_00 0397	CY BB	cytochrome b-245, beta polypeptide (chronic granulomatous disease); synonyms: CGD, NOX2, GP91-1, GP91PHOX, GP91-PHOX; Homo sapiens cytochrome b-245, beta polypeptide (chronic granulomatous disease) (CYBB), mRNA.;

		Cytochrome b-245, beta polypeptide (chronic granulomatous disease)
BC004 449	ME F2B	synonym: RSRFR2; Homo sapiens MADS box transcription enhancer factor 2, polypeptide B (myocyte enhancer factor 2B), mRNA (cDNA clone IMAGE:2819918), complete cds.; Myocyte enhancer factor 2B
NM_14 7780	CT SB	cathepsin B; synonyms: APPS, CPSB; APP secretase; preprocathepsin B; cathepsin B1; amyloid precursor protein secretase; cysteine protease; Homo sapiens cathepsin B (CTSB), transcript variant 2, mRNA.; Cathepsin B
NM_08 0821	C2 0orf108	chromosome 20 open reading frame 108; synonyms: 5A3, dJ1167H4.1, DKFZP434A1114; Homo sapiens chromosome 20 open reading frame 108 (C20orf108), mRNA.; Chromosome 20 open reading frame 108
NM_03 3089	ZC CHC3	zinc finger, CCHC domain containing 3; synonyms: C20orf99, MGC104290; Homo sapiens zinc finger, CCHC domain containing 3 (ZCCHC3), mRNA.; Zinc finger, CCHC domain containing 3
NM_01 8196	TM LHE	trimethyllysine hydroxylase, epsilon; synonyms: TMLH, BBOX2, XAP130, FLJ10727; epsilon-trimethyllysine 2-oxoglutarate; epsilon-trimethyllysine hydroxylase; butyrobetaine (gamma), 2-oxoglutarate dioxygenase (gamma-butyrobetaine hydroxylase) 2; Homo sapiens trimethyllysine hydroxylase, epsilon (TMLHE), mRNA.; Trimethyllysine hydroxylase, epsilon
NM_03 1216	SE H1L	SEH1-like (S. cerevisiae); synonyms: Seh1, SEH1A, SEH1B, SEC13L; isoform 2 is encoded by transcript variant 2; sec13-like protein; nucleoporin Seh1; sec13 like protein; Homo sapiens SEH1-like (S. cerevisiae) (SEH1L), transcript variant 2, mRNA.
CR627 133		Homo sapiens mRNA; cDNA DKFZp686J154 (from clone DKFZp686J154).; Hypothetical protein LOC342892
AK0975 17		unnamed protein product; Homo sapiens cDNA FLJ40198 fis, clone TESTI2019975, weakly similar to TRICHOHYALIN.; FLJ40198 protein
AV7360		AV736006 CB Homo sapiens cDNA clone

06		CBFBEA05 5', mRNA sequence.; G protein-coupled receptor 89A
NM_020375	C12orf5	chromosome 12 open reading frame 5; synonym: TIGAR; transactivated by NS3TP2 protein; Homo sapiens chromosome 12 open reading frame 5 (C12orf5), mRNA.; Chromosome 12 open reading frame 5
THC2613527		
A_24_P620521		
NM_006241	PP1R2	protein phosphatase 1, regulatory (inhibitor) subunit 2; synonyms: IPP2, MGC87148; Homo sapiens protein phosphatase 1, regulatory (inhibitor) subunit 2 (PPP1R2), mRNA.; Protein phosphatase 1, regulatory (inhibitor) subunit 2
A_24_P636974		
NM_000692	ALDH1B1	aldehyde dehydrogenase 1 family, member B1; synonyms: ALDH5, ALDHX, MGC2230; aldehyde dehydrogenase 5; mitochondrial aldehyde dehydrogenase X; acetaldehyde dehydrogenase 5; ALDH class 2; Homo sapiens aldehyde dehydrogenase 1 family, member B1 (ALDH1B1), nuclear gene encoding mitochondrial protein, mRNA.; Aldehyde dehydrogenase 1 family, member B1
BC043153	CTAGE6	Homo sapiens CTAGE family, member 6, mRNA (cDNA clone MGC:41943 IMAGE:5296763), complete cds.
A_24_P689479		
A_24_P790361		
NM_021033	RAP2A	RAP2A, member of RAS oncogene family; synonyms: KREV, RAP2, K-REV, RbBP-30; RAP2, member of RAS oncogene family (K-rev); Homo sapiens RAP2A, member of RAS oncogene family (RAP2A), mRNA.
NM_207038	TCF12	transcription factor 12 (HTF4, helix-loop-helix transcription factors 4); synonyms: HEB, HTF4, HsT17266; isoform b is encoded by

		transcript variant 4; Homo sapiens transcription factor 12 (HTF4, helix-loop-helix transcription factors 4) (TCF12), transcript variant 4, mRNA.; Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)
A_24_P 843552		
THC25 47775		
BC039 021		Homo sapiens cDNA clone IMAGE:6043059, partial cds.; CDNA clone IMAGE:6043059
NM_00 4232	SO CS6	suppressor of cytokine signaling 6; synonyms: CIS4, SSI4, SOCS4, STAI4, STATI4, HSPC060; STAT induced STAT inhibitor-4; suppressor of cytokine signaling 4; cytokine- inducible SH2 protein 4; Homo sapiens suppressor of cytokine signaling 6 (SOCS6), mRNA.; Suppressor of cytokine signaling 6
A_24_P 928250		
THC25 85049		Q28EK6_XENTR (Q28EK6) CHK1 checkpoint homolog (S. pombe), partial (10%) [THC2585049]
AW389 851		RC2-ST0171-271099-031-d05 ST0171 Homo sapiens cDNA, mRNA sequence.; Family with sequence similarity 38, member A
AK0254 31		Homo sapiens cDNA: FLJ21778 fis, clone HEP00201.; receptor expressed in lymphoid tissues like 1; Receptor expressed in lymphoid tissues like 1
BC044 628		Homo sapiens mRNA similar to collagen, type XVII, alpha 1 (cDNA clone IMAGE:5431129).
A_32_P 121674		
ENST0 0000343905		PREDICTED: Homo sapiens ankyrin repeat domain 58 (ANKRD58), mRNA [XM_293380]
AA7784 30		zf39f12.s1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:379343 3', mRNA sequence.; ATP synthase, H+ transporting, mitochondrial F0 complex, subunit s (factor B)
A_32_P		



124521		
THC25 51769		AA411302 zv24g06.r1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:754618 5', mRNA sequence [AA411302]
NM_00 2582	PA RN	poly(A)-specific ribonuclease (deadenylation nuclease); synonym: DAN; Homo sapiens poly(A)- specific ribonuclease (deadenylation nuclease) (PARN), mRNA.; Poly(A)-specific ribonuclease (deadenylation nuclease)
NM_19 8935	SS 18L1	synovial sarcoma translocation gene on chromosome 18-like 1; synonyms: CREST, LP2261, KIAA0693, MGC26711, MGC78386; calcium-responsive transactivator; SYT homolog 1; Homo sapiens synovial sarcoma translocation gene on chromosome 18-like 1 (SS18L1), mRNA.; Synovial sarcoma translocation gene on chromosome 18-like 1
ENST0 0000316294		Seven transmembrane helix receptor. [Source:Uniprot/SPTREMBL;Acc:Q8NHC0] [ENST00000316294]
THC25 30832		LAMB1_HUMAN (P07942) Laminin beta-1 chain precursor (Laminin B1 chain), partial (4%) [THC2530832]
THC27 52542		
ENST0 0000295549		Homo sapiens hypothetical gene supported by BC013438, mRNA (cDNA clone IMAGE:3899073), partial cds. [BC013438]
NM_01 5254	KIF 13B	kinesin family member 13B; synonyms: GAKIN, KIAA0639; guanylate kinase associated kinesin; kinesin 13B; Homo sapiens kinesin family member 13B (KIF13B), mRNA.; Kinesin family member 13B
CR744 556		CR744556 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGp998M072299 ; IMAGE:925230 5', mRNA sequence.
NM_13 8356	SH F	Src homology 2 domain containing F; synonyms: MGC126650, MGC126652; Homo sapiens Src homology 2 domain containing F (SHF), mRNA.
THC26 80668		BU678941 UI-CF-DU1-aat-a-09-0-UI.s1 UI- CF-DU1 Homo sapiens cDNA clone UI-CF-DU1-



		aat-a-09-0-UI 3', mRNA sequence [BU678941]
THC27 18727		
AA7196 41		zg52f01.s1 Soares_pineal_gland_N3HPG Homo sapiens cDNA clone IMAGE:396985 3', mRNA sequence.; Neurotrimin
A_32_P 202621		
THC27 46246		Q9TD22_9SMEG (Q9TD22) NADH dehydrogenase subunit II, partial (5%) [THC2746246]
AK0936 41	FA M120AOS	Homo sapiens cDNA FLJ36322 fis, clone THYMU2005507.; family with sequence similarity 120A opposite strand; Family with sequence similarity 120A opposite strand
THC26 46885		
THC25 64099		BC005134 RPL14 protein {Homo sapiens} (exp=-1; wgp=0; cg=0), partial (82%) [THC2564099]
A_32_P 23154		
NM_01 2430	SE C22A	SEC22 vesicle trafficking protein homolog A (S. cerevisiae); synonym: SEC22L2; sec22 homolog; SEC22 vesicle trafficking protein-like 2; Homo sapiens SEC22 vesicle trafficking protein homolog A (S. cerevisiae) (SEC22A), mRNA.
NM_00 1023567	GO LGA8B	golgi autoantigen, golgin subfamily a, 8B; synonyms: GOLGA5, KIAA0855; Homo sapiens golgi autoantigen, golgin subfamily a, 8B (GOLGA8B), mRNA.; Golgi autoantigen, golgin subfamily a, 8A
A_32_P 32923		
A_32_P 36412		
NM_01 5122	FC HO1	FCH domain only 1; synonym: KIAA0290; Homo sapiens FCH domain only 1 (FCHO1), mRNA.
BC017 654		Homo sapiens cDNA clone IMAGE:3887603.; hypothetical gene supported by AK001829; Hypothetical gene supported by

		AK001829
AV7145 56		AV714556 DCB Homo sapiens cDNA clone DCBADG06 5', mRNA sequence.; Family with sequence similarity 91, member A1
NM_00 6313	US P15	ubiquitin specific peptidase 15; synonyms: UNPH4, KIAA0529, MGC74854, MGC131982, MGC149838; deubiquitinating enzyme; Homo sapiens ubiquitin specific peptidase 15 (USP15), mRNA.; Ubiquitin specific peptidase 15